American Statistical Association

Organized November 27, 1839
Incorporated 1841

The American Statistical Association is a scientific and educational organization. Its membership is not confined to professional statisticians but includes economists, business executives, research directors, government officials, university professors, and other persons who are seriously interested in the application of statistical methods to practical problems, in the development of more useful methods, and in the improvement of basic statistical data. Engineers, mathematicians, biologists, actuaries, sociologists, psychologists, and representatives of many other professions are included in the membership of the Association.

Regular membership	$ 5.00*
Introductory membership (for the first dues payment of applicants under 30 years of age)	3.00
Members subscription to *Biometrics Bulletin*	1.00
Associate membership in the Biometrics Section	2.00
Contributing membership	25.00

Additional information about the Association and membership application forms may be secured from the Secretary, Lester S. Kellogg, 1603 K Street, N.W., Washington 6, D.C.

Chapters of the Association have been organized in important cities in the United States and in Havana, Cuba. Information concerning the meetings of these Chapters may be secured from the District Representatives listed on another page.

The Editors welcome the submission of articles and notes for possible publication in the JOURNAL. Manuscripts should be sent to the Editor, Journal of the American Statistical Association, 1603 K Street, N. W., Washington 6, D.C. (1) Send two or more copies of your manuscript (preferably in separate envelopes, to avoid danger of loss). (2) Leave one half of the first page of your manuscript blank, to be used for instructions to the printer. (3) Attach to the front of your manuscript a fifty word outline or summary. Authors who wish suggestions about the preparation of manuscripts and charts or information about editorial policies should address their inquiries to the Editor. Book reviews should be sent to Major Oscar K. Buros, 301 South Courthouse Road, Arlington, Virginia.

Subscription rate, $6.00 per annum. Single numbers of volumes 23 to 36, $1.50 per copy postpaid, except Proceedings Supplements to volumes 23 to 30, $2.00 per copy postpaid. Prices for sets and discounts for orders in quantity available on request.

* $5.00 of which is for one year's subscription to the Journal of the American Statistical Association, and 50 cents for one year's subscription to the American Statistical Association Bulletin.

Published Quarterly by the AMERICAN STATISTICAL ASSOCIATION
Publication Office: 450 Ahnaip Street, Menasha, Wisconsin. Editorial Office: 1603 K Street, N.W., Washington 6, D.C. Acceptance for mailing at special rate of postage provided for in the Act of February 28, 1925, embodied in paragraph 4, section 538, P. L. & R., authorized March 26, 1936. Entered as second class matter at the post office at Menasha, Wisconsin.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Volume 41 | March, 1946 | Number 233

ARTICLES

PROCEEDINGS OF THE 105TH ANNUAL MEETING

BOOK REVIEWS

Index to Volumes 1–34, 1888–1939, may be obtained from the ASA. The JOURNAL is also indexed in the Industrial Arts Index, the International Index to Periodicals, and the Public Affairs Information Service Bulletin.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Volume 41 | June, 1946 | Number 234

ARTICLES

BOOK REVIEWS

Index to Volumes 1–34, 1888–1939, may be obtained from the ASA. The JOURNAL is also indexed in the Industrial Arts Index, the International Index to Periodicals, and the Public Affairs Information Service Bulletin.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Volume 41 September, 1946 Number 235

ARTICLES

BOOK REVIEWS

Index to Volumes 1–34, 1888–1939, may be obtained from the ASA. The JOURNAL is also indexed in the Industrial Arts Index, the International Index to Periodicals, and the Public Affairs Information Service Bulletin.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Volume 41 December 1946 Number 236

ARTICLES

BOOK REVIEWS

Index to Volumes 1–34, 1888–1939, may be obtained from the ASA. The JOURNAL is also indexed in the Industrial Arts Index, the International Index to Periodicals, and the Public Affairs Information Service Bulletin.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 233 MARCH 1946 *Volume 41*

THE ADVANCING STATISTICAL FRONT*

WALTER A. SHEWHART
Bell Telephone Laboratories

THE ATOMIC BOMBS on Hiroshima and Nagasaki have made the world conscious, as perhaps no other event has done, of the potential significance of applied science. This war revealed to all that the co-operation of scientists on a gigantic scale literally revolutionized the method of carrying on wars. By the grace of God, the nations of the world have been given another chance to live in peace—another chance to explore the endless frontier of science in search of new ways of constructively serving mankind.

Most people will agree that the problems of war are inherently simpler than those of attaining the full advantages of science in a world at peace. For war is a simple business in the sense that the objectives are clear and heartily supported by an overwhelming majority of the citizens of a nation; whereas peace-time problems do not have the same dramatic intensity, immediacy, and urgency to each citizen. However, the atomic bomb has brought people to appreciate the importance of the *end* to which science is applied. All shudder at the thought of the destructiveness of another war; all like to dream of the world of tomorrow in which the object of applied science is the satisfaction of human wants. However, these wants cannot be scientifically determined without appeal to the sampling procedures of the statistician. To the citizen of a democracy there is certainly some dramatic significance in the fact that the statistician must be called in to sample human wants in order that science may be applied effectively in peace.

It seems fitting that, at this first annual meeting after the catastrophic war through which we have just passed, our attention should be directed toward the role that statistics can play in the years ahead in

* Address as retiring president of the American Statistical Association, Cleveland, Ohio, January 26, 1946.

making science *serve* humanity and toward the role that statisticians can play in giving each individual some appreciation of the dramatic intensity, immediacy, and urgency to him of some of the problems that the statistician alone can solve. With this thought in mind, I have chosen to speak on the *advancing front of statistics in its service*, directly and through other scientists, to the members of society as *individuals*.

If statistics can be made to minister to *all* the people, if the majority of the people can be made to realize this contribution of statistics, if those who practice statistics follow a strict code of conduct, if statisticians assume their responsibility for the proper training of neophytes before turning them loose as statisticians, then the future of the statistical profession is assured.

In the established sense of the term "profession" as applied to law, medicine, and religion, there is an implication of a devotion to a higher purpose than that of personal profit or the earning of a livelihood. Tonight I wish to keynote ministering to the people at large as an essential requirement of any profession. At some time or other between the cradle and the grave, every one of us calls upon members of the medical, legal, and religious professions. In so far as these serve us well, we come to appreciate them as members of their profession because we recognize that through long years of study and training they have placed themselves in a position to do for us what we cannot do for ourselves. If statistics can be made to serve each of us in a way that we individually recognize and appreciate, statisticians will no longer be referred to in some quarters as the showgirls of the market place because they live by their figures alone.

The high purpose of a diverse group can be built up and maintained only by an effective organization. By getting a perspective of the advancing research front and the vast armies of pure and applied scientists with whom we must cooperate if we are to make our greatest contribution to the service of science, we can come to realize fully the need for a strong professional statistical organization.

THREE TYPES OF SCIENTIFIC RESEARCH

As a background for what follows, it is essential that we distinguish between three broad and somewhat overlapping categories of research. In the recent report,[1] "Science, the Endless Frontier," by Vannevar Bush, Director of OSRD, to the President of the United States on a

[1] United States Government Printing Office, Washington, 1945.

program for post-war scientific research, the following three categories of research are recognized and briefly described: pure research, background research, and applied research and development.

Pure research is there defined as research without specific practical ends. It is directed toward the acquisition of knowledge and the understanding of nature and its laws. *Background* research is typified by the collection of economic, business, census, meteorological, and other types of data; the determination of physical and chemical constants; and the establishment of standards of quality for foods, drugs, and other manufactured materials. *Applied* research is generally directed toward some practical end determined by human wants and it is usually expected that the cost of such research will be met through the benefits that accrue from the research itself.

THE STATISTICAL FRONT IN THE THREE FIELDS OF RESEARCH

Pure Research. Since the turn of the century, great progress has been made in extending the statistical front in pure research. Advances have been made in the field of the statistical description of stellar phenomena by Chandrasekhar and others; in the development of statistical mechanics by Einstein, Bose, Fermi, Dirac and others; in the statistical description of meteorological phenomena; and in the development of statistical theories in genetics and many other fields. Statistics has progressed also in psychometrics and econometrics. But perhaps the greatest progress within the past two decades has been in the field of mathematical statistics along the lines typified by the abstract material published in the *Annals of Mathematical Statistics*.

Such work has been highly individualistic. Much of it has been done by those unknown to most of the members of the American Statistical Association. In fact, many of those who have done most to extend the frontiers of mathematical statistics are not known officially as statisticians: Even R. A. Fisher, Hotelling, and Wilks are not listed as statisticians in their college registers. Likewise the men who have contributed to the knowledge of nature through the application of statistics are seldom known to their fellow scientists as statisticians but rather as professional men in the subject matter field wherein they contributed. Statistics, in other words, has no particular subject matter field solely unto itself.

Moreover, most of the contributions of statistics to the different subject matter fields are mathematical in character; and in so far as such contributions are mathematical, it is little wonder that fellow scientists often question the existence of a statistical front as such.

Many think simply of the different subject matter fronts while others think of statistics as a part of the mathematical front.

Let us pause to examine the significance of this situation from the viewpoint of professional organizations in the field of statistics. To maintain an advancing front in *pure* statistical research both in the subject matter fields and in mathematics requires support as does pure research in every field. It is perhaps true that each subject matter field, in so far as pure research is concerned, can obtain support for the development of the statistical aspects of nature essential to that field. Those engaged in pure research in the separate fields will undoubtedly recognize the need for courting the mathematical statistician engaged in extending the boundaries of mathematical statistics per se, for perhaps never before in the history of science have scientists in the subject matter fields appreciated so wholeheartedly the contributions of mathematics. In fact, scientists are coming more and more to realize that most, if not all, theories of nature are mathematical in character at least to the extent that the idealized abstract models of natural phenomena are described in mathematical terms. There is, of course, a need for the mathematical statistician to come into contact with the problems revealed by nature in the separate sciences if mathematics is to contribute to the solution of such problems.

Those working to extend the boundaries of mathematical statistics need to have an organization such, for example, as the Institute of Mathematical Statistics to serve as a go-between for pure mathematics on the one hand and *pure science* on the other. To their fellow researchers in other fields of pure science men working in this realm of statistics will probably always be known primarily as mathematicians. Certainly to those working in the fields of background and applied science and to the public, such statistical research workers will probably always be known as mathematicians. Hence, so long as we think of the contributions of statistics to pure research, we may question the need for an organization beyond that of mathematical statisticians. When, however, we consider the service to the public of pure statistical research through its contributions to background and applied research, we find that an organization representing the common interests of all statisticians is a necessity. Otherwise we cannot fulfill our obligations to the public upon which we must depend in the end for recognition and support.

Background and Applied Research. Turning to the fields of background and applied research, three points should be kept in mind: 1) such research must involve the *cooperation* of many individual ap-

plied scientists, sometimes literally hundreds, 2) such research is directed toward obtaining knowledge useful to many individuals and sometimes to practically the whole public, and 3) the usefulness of such work must be recognized by those who receive the benefits in order to gain their financial and other support: applied research must pay its own way.

What I want to do is to indicate in some detail the endless field open to the statistician in applied research and to indicate briefly some of the problems involved in obtaining the cooperation of individual applied scientists and of groups of applied scientists in carrying out a few major lines of application that are of significance to the public.

First, let us consider simply the background research of standardization as carried on by groups like the American Standards Association and the American Society for Testing Materials. Work on standardization projects often requires the cooperation of many different kinds of scientists representing the public, industry, and government: the number of such projects runs into thousands. Each such standard is, in general, arrived at through committee action often extending over several years: many compromises must be made, some because of a lack of scientific knowledge, some because of economic requirements, and some because of other limitations. Each standard must be, as it were, a tailor-made job that is satisfactory to different group interests. Most of the applied scientists cooperating in such a project are not familiar with statistical theory and application; many of them are, in fact, antagonistic to the statistician. Yet these scientists are in the saddle and decide what shall be done.

Even if we assume that we have the ideal condition in which a standardizing committee is completely sold on the applications of statistics, there remains the problem of translating the contributions of the statistician into the written specification in a way that will be understood by those who use it—a group that often includes hundreds of people; indeed in the case of a consumer standard, thousands of people. The problem of putting statistics to work in standardization, one of the fields of background research touching all of us as consumers of foods, drugs, and other manufactured goods, is a tremendous job calling for the cooperation of applied scientists working not only as individuals but as groups. *This is not a problem that can be solved by statisticians working independently:* yet in its solution there is room for hundreds of scientists with basic training in statistics.

In applied research, let us consider only industrial research. In 1942 there were 2254 industrial research laboratories: today there are many

more. Yet I doubt that many of the directors of these laboratories have any appreciation of the potential contributions of statistics. In 1939 there were 184,230 manufacturing firms in the United States: today there are perhaps as many or more. Yet I doubt that management in any but a few of these firms is aware of the potential contributions of statistics. Of course, many employ statistical clerks and a few may even employ one or more persons called statisticians but their statistics is usually limited to business and economics. There is a big gap to be filled between the contributions of the mathematical statistician in the Annals, for example, and the appreciation of the value thereof on the part of management in any large percentage of industrial laboratories and manufacturing firms. Nevertheless, statisticians have a great opportunity in this almost unlimited field to serve the public at large as users of the manufactured goods turned out.

The contribution of statistics in these two fields of applied research lies not so much in solving the problems usually put to the statistician by those not statistically trained as in helping to coordinate the steps of specification, production, and inspection. This may be considered as a scientific experiment for making the most efficient use of human effort in the production of goods to satisfy human wants. The long range contribution of statistics depends not only upon getting many highly trained statisticians into industry but also upon creating a *statistically minded new generation of those physicists, chemists, engineers, and others* who will develop and direct the mass production processes of tomorrow.

Let us look at the problem of interesting the members of professional organizations in the three fields of pure, background, and applied research. Broadly speaking, professional scientific organizations may be classified under these three headings although the boundaries are not distinct. At the pure research level, we have those organizations representing the basic natural and social sciences; at the level of applied research, we have large professional groups such as the engineering societies, several of which have memberships of 20,000 or more; at the level of background research we have the American Society for Testing Materials, The American Standards Association, and certain trade associations. By and large, these professional groups with a combined membership of more than a hundred thousand are responsible for the development of the scientific research front. These groups in the end will determine to what extent, if any, statistical methods will be used in their work. We must first get the membership of these scientific groups to give us a chance to show how statistics can help. Then we must work in close cooperation with these groups in putting statistics to work for

the public good. Statisticians must assume a large part of the responsibility for finding out how they can best serve the thousands of scientists in the different subject matter fields and then they must cooperate with other professional groups in the different subject matter fields to see that statistical method is applied wherever it will yield useful results.

THREE WAYS OF SERVING FELLOW SCIENTISTS

Three ways of serving fellow scientists are through joint meetings, through joint committees, and through publications. For years our Association has been accustomed to holding joint meetings with other groups, particularly in the social sciences. More recently joint meetings have been held with the Institute of Mathematical Statistics and the Biometrics Section has held many joint programs with groups having interests similar to theirs. By and large, however, there have been no joint meetings with the majority of the scientific groups primarily responsible for pure research in other than social science fields nor with the many large groups responsible for background and applied research.

A joint committee type of approach to applied research in engineering was started in 1929. In that year E. B. Wilson, then President of the American Statistical Association, cooperated with a small group of statisticians in getting the interest of certain professional engineering groups in setting up what was and is called the Joint Committee for the Development of Statistical Applications in Engineering and Manufacturing. This committee was actively sponsored by the American Statistical Association, the American Society for Testing Materials, The American Society of Mechanical Engineers, and the American Mathematical Society. The membership was made up of two representatives from each of these societies. This committee, working in cooperation with the officers of the sponsor societies, has been largely influential in initiating most of the activities that have taken place since then in organizing statistical effort in the engineering field.

It was also realized at that time by the same group of statisticians that it was desirable to launch a committee interested in the application of statistics to the background research of standardization. Such a committee was set up and has been active ever since in the American Society for Testing Materials. A short time before the war another small group of statisticians recognized the desirability of getting formal action on the application of statistics in quality control through the standardization of simplified methods. With the cooperation of the War Department and the American Standards Association a committee

was set up under the auspices of the Amerian Standards Association. This committee turned out three emergency war standards in statistical quality control and these were used extensively not only in America but also in Australia, Canada, and England.

During the early years of the war, there were often discussions by the Board of the American Statistical Association about the need for a joint committee representing some of the major fields of war research, more nearly on the pure research level, to act in an advisory capacity on the applications of statistics. Nothing was done officially by the Association, but statisticians working independently in cooperation with the officers of the National Research Council and particularly with Dean Eisenhart, chairman of the Division of Physical Sciences, organized in 1943 a committee on applied mathematical statistics with subdivisions on physics, chemistry, psychology, economics, mathematical statistics, biology, and engineering. This committee has had from the beginning liaison members with the War and Navy Departments and the War Production Board. More recently, liaison representation has been secured on the part of the National Physical Laboratory of Great Britain and the Council for Scientific and Industrial Research of Australia. The latest development in this organized committee effort is the establishment of a committee, under the joint direction of the National Research Council and the Social Science Research Council, charged with the study of the scientific problems of measuring public opinion, attitudes, and consumer wants.

Much of the work of these committees has been of an advisory nature. For example, the joint committee on engineering, which is now sponsored not only by the original sponsors but also by the Institute of Mathematical Statistics and the American Institute of Electrical Engineers, has cooperated with different engineering groups in setting up special committees on the application of statistics within these groups and in arranging for the presentation of papers on applied statistics on the programs of many of the engineering societies. Just recently such a committee was set up in the American Institute of Electrical Engineers.

Some of these committees have operated to secure not only national cooperation but also international. For example, the Joint Committee, through its sponsors, was influential in starting the movement in Great Britain that led to the organization of a committee under the auspices of the British Standards Institution. Interest thus aroused has been credited by our English friends as the spark that started the movement which led to the formation of the Agricultural and Industrial Research Section of the Royal Statistical Society.

To date, however, such effort has been largely individualistic even perhaps to the extent that few statisticians in either the American Statistical Association or the Institute of Mathematical Statistics are aware of it. Furthermore, these committees have only scratched the surface of the work that needs to be done by way of cooperation with other scientific groups. Perhaps they set a pattern, however, that deserves more study on the part of those interested in the statistical profession. In passing, it may be remarked that in setting up such joint committees it seems desirable that at least two representatives be chosen from each participating group. One should be an "elder statesman" in the subject matter field of the group, in addition to being a person who is broad-minded and interested in all new developments that might contribute to the work of his profession. At the same time he should be a sympathetic but hard-boiled critic. The other representative should not only be qualified in the particular subject matter field but also in the technical applications of statistics in that field.

Let us turn now to the problem of serving our fellow scientists by means of publications. At least two media must be considered: journals of the many different scientific organizations, and a thoroughly up-to-date statistical journal containing methodological and subject matter articles of general interest to statisticians working in the three fields of research. To date, statisticians working as individuals have published articles in many professional journals, and the editorial board of the *Journal of the American Statistical Association* plans to provide a journal of high calibre that will carry material of broad general interest. For example, the article by Wald in the September issue of the *JASA* is of potential interest to members of many professional organizations in the field of pure and applied science. Hence progress is being made. However, we still have a long way to go in developing a well-thought-out plan for reaching the thousands of scientists in different subject matter fields.

THREE WAYS OF SERVING THE PUBLIC

Let us now consider some of the ways in which statisticians serve the public. As already noted, it is essential for the existence of a strong professional organization that the public recognize to some extent the service rendered by its members. If, for example, the public becomes aware that by extending the use of statistics in many fields of research the potential contributions of that research in satisfying human wants may be maximized, it will not only appreciate the statistician but it may also bring pressure to bear wherever necessary to see that desirable applications of statistics are made.

Since, as already noted, the professional scientific groups in the separate subject matter fields mainly determine the extent to which statistics will be used, I shall use the three-fold classification of contributions of research that are essential to the public proposed in Dr. Bush's report. These are in the war against disease, in the maintenance of national security, and in increasing public welfare. By adopting this classification we may more easily see how the need for the contribution of the statistician fits into the picture of the need for scientific progress presented in this report and now being discussed so extensively in scientific literature.

War Against Disease. The study of vital statistics, including the study of epidemics and the causes of disease, was one of the earliest fields to which statisticians contributed. Statistical studies to determine hereditary susceptibility to disease have long been of interest from the viewpoint of preventive medicine. The determination of the effects of drugs and medical treatments is inherently a very complicated sampling process. Statistical methods, of course, have been long used in studying the effects of specific drugs in the laboratory. Here statistics plays an important part not only in the interpretation of results but also in the design of efficient experiments. Statistics also plays an important role in the study of the effects of different kinds of bacteria, climatic conditions, and the like, on health. It should not be hard to get across, even at the high school level, some appreciation of such uses of statistics. In the whole field of foods and dietetics the public could be made to appreciate something of the importance of statistics.

National Security. For years government agencies have employed most of the statisticians. Their services have long been recognized as of great importance in the collection and assembly of factual information about the supply and demand of labor as well as the supply and demand of manufactured goods and agricultural products. The statistician has for years played an important part in agricultural experimentation, in the study of social trends and in the general field of economics. In so doing, he has made available information of great value to those responsible for planning national security.

However, at the beginning of the present war there was little appreciation of the value of the statistician in government other than in the fields mentioned above. Even in setting up the National Roster, the profession of statistics was given little recognition. However, during the war this situation was greatly modified. Within such agencies as the War and Navy Departments, the War Production Board, the Office of the Surgeon General, and the National Defense Research Committee,

statisticians were given a chance to make important contributions. These extended beyond the business and social fields into the fields of background and applied research in the natural sciences. If and when the record can be made public, we will find a surprisingly large contribution to the war effort in the fields of personnel placement, operational research, development of improved war materials, and the control of the quality of much of the material used by the Armed Forces.

However, the organization of such statistical effort depended primarily upon the initiative of small groups of statisticians. Had we had at the beginning of the war, a strong professional organization interested in all fields of applied research, I think that our contribution would have been even greater. Prior to the war statisticians had not given sufficient attention to the developing front of statistics along new lines, particularly in the field of applied natural science. Organized planning was so inadequate that there were not enough statisticians qualified even to teach the large numbers of technicians required.

From the viewpoint of future security, a strong professional organization is needed to keep building sufficient public support for the many *new* fields of application that are likely to be of use in developing our national security. Moreover, if the potential contributions of statistics in this field are to be realized, much remains to be done in getting across to a large percentage of the public the part played by statistics in this war. Professional organizations in other fields of applied science have already taken steps to get recognition for their work. In the American Statistical Association, the officers and many of the members have given much consideration to this problem. One of the steps being considered is the publication of technical articles in the *JASA* to acquaint all statisticians with war-time contributions. This, however, is not enough. We need to reach great numbers of people who never see the *JASA*. Through short articles and editorials in the press and popular magazines, an effort should be made to indicate in an understandable manner not only the contributions that statistics made in the war but also the necessity of continually developing such applications as a part of any plan for security.

Public Welfare. In a democracy, public opinion and attitudes play an important role. In so far as public policies are determined upon the basis of public opinion, it is certainly of interest to everyone that such opinion be measured *in the best possible way*. All of you are well aware of the complicated statistical sampling problems as well as the psychological, economic, and other problems involved in the measurement of public opinion. However, it would not be extremely difficult, I believe,

to make most people have an appreciation for some of the more elementary statistical problems involved.

Even a group of men sitting around an old pot-belly stove in a backwoods general store could be made to see the fallacy in the *Literary Digest* Poll. If the public can be convinced of the importance of using adequate statistical care in such work, is it not reasonable that they would support a professional organization that would assume the responsibility for providing adequate statistical techniques for those working in this field and for acquainting the public with information that will help individuals to form valid judgments on questions involving public opinion. It seems to me that this is one of the public services that the Association can render.

Another example of the usefulness of statistics to all individuals is the establishment of consumer standards of quality. One aspect of this problem of particular interest is that it requires the cooperation not only of applied natural but also of applied social scientists. Much work, of course, has been done by members of the Marketing Research Association, by certain government agencies, and by standardizing groups. Success in this field is, of course, closely tied up with success in solving the problems involved in measuring public opinion and attitudes. So far, a large part of the work in consumer standards has been in determining preferences between things known to the consumer. In many cases, more needs to be done by way of determining what *might* sell, and what might give even greater satisfaction to customers. Here I am thinking primarily of what might be done at the applied research level by a team composed of physicists, chemists, psychologists, economists, market research men, statisticians, and possibly physiologists in solving some of the problems of establishing consumer standards. During the past two years, these problems have been discussed in several issues of the *Bulletin* of the American Society for Testing Materials and the American Standards Association publication.

To make further progress it is essential that statisticians, familiar with what is known at the present time about the science of measuring opinion and attitudes, cooperate with professional groups like the American Standards Association and the American Society for Testing Materials to establish standards in terms of physical, chemical, and other properties. More than this, it would be highly desirable if we could find some objective physiological measure of human reactions that would be more reliable in many instances than verbal expression.

The problem of standardization is difficult even when considered in

terms of theoretically measurable characteristics of a product: when we couple with this the problem of determining in a quantitative manner the correlation between human reaction and physically measurable characteristics of manufactured goods, it is obvious that the problems are vastly more complicated. All of these problems are inherently statistical and unless statisticians as a professional group assume their responsibility in helping to solve them, statistics will have lost one of its great opportunities to serve the public in a way that the public can understand. If, on the other hand, these problems are attacked in a half-hearted manner and are muffed, the day of full public recognition *and appreciation of the statistician will be greatly delayed. In other* words, we have here a frontier for applied statistical research of the highest order. We have here a problem in which the public is greatly interested not only as individuals but as organized groups—we have here a problem that the public *wants* solved.

In closing this section on public welfare, let us consider briefly the question: What can statisticians contribute to general education? The recent report on *General Education in a Free Society*, published by Harvard University, gives the following definition of general education: "It (general education) is used to indicate that part of a citizen's whole education which looks first of all to his life as a responsible human being and citizen; while the term 'special education' indicates that part which looks to the citizen's competence in some vocation."

I have recently read not only the Harvard report but many other discussions on general education and I am pleased to find that several emphasize the importance of some knowledge of scientific method as a part of general education. In a few cases, the authors mention statistics but not in a way that indicates any relationship to scientific method. This is not surprising because there has been little effort in statistical teaching to emphasize how statistical method is related to the three steps in scientific method, namely, hypothesis, experiment, and test of hypothesis. If the student were made to appreciate that, because of the statistical nature of observable phenomena, it is often necessary to replace the three steps by statistical hypotheses, statistically designed experiments, and statistical tests of hypotheses, he would appreciate the developing growth of scientific method and the contribution of statistics thereto.

From the viewpoint of general education, statistics is *not* simply a tool as is so often stated but *a scientific way of looking at the universe:* statistical method is not something apart from scientific method but *is* scientific method adjusted to the generally accepted view that science

is not exact but only probable. Statistics is like the well-dressed bride in that statistics "wears":

Something old: scientific method.

Something new: *improvements on each of the three steps in scientific method.*

Something borrowed: Mathematics and logic.

Something blue: Contributions in the "blue chip" class in every subject matter field.

CONCLUDING REMARKS

The statistical front is the whole field of science. Outposts of that front exist in every subject matter field of research. Many outposts are manned by small scouting parties of pioneering souls; others are well-organized groups like the Biometrics Section of the American Statistical Association, the Econometric and Psychometric Societies, the Institute of Mathematical Statistics, and the American Marketing Association. Some of these outposts have already advanced out of sight of the others; in no instance are the lines of communication adequate—even worse, there is no available map showing the location, personnel, and objectives of these outposts.

Picture these outposts on the scientific frontier each with a mission to perform that is of vital interest to the public, but without any *one* headquarters statistical organization to which appeals can be directed for the latest statistical results from the similar outposts in other fields of pure, background, and applied research; without any one headquarters organization to stimulate the training of new recruits in the separate subject matter fields; without any one headquarters organization to maintain adequate public support for their work vital to our national security, war against disease, and public welfare.

To insure continued advancement on the statistical front, we need a strong professional organization backed by *all* statisticians working in the many different subject matter fields: an organization whose purpose shall be to foster in the broadest manner the *science* of statistics and its applications; and to promote a unified, effective, organized effort in advancing the common interests of all professional statisticians.

To this end, such a professional organization needs to conduct joint meetings and conferences with many organizations in the fields of pure, background, and applied research with whom official contact has never been made in the past; to establish joint committees with other scientific groups wherever necessary; to produce and distribute publications

devoted to statistical methodology and statistical knowledge of common interest to all professional statisticians; to provide the public through the press, lecture platform, and radio with reliable information about the science of statistics and its contributions; and to stimulate statistical research and promote high professional standards in all statistical work.

I hope and believe that the American Statistical Association with the cooperation of other organized societies interested in statistics, can be developed into a strong organization to serve the common interests of *all* professional statisticians in these groups.

ON THE DESIGN OF A SAMPLE FOR DEALERS' INVENTORIES

W. EDWARDS DEMING
Bureau of the Budget
WILLARD SIMMONS*
Office of Price Administration

I. STATEMENT OF THE PROBLEM

ADMINISTRATION of the tire-rationing program required current and reliable knowledge of the numbers of passenger, truck, and other civilian-grade tires held by dealers, mass distributors, and manufacturers. To fill this need, inventories of tire dealers were taken each quarter, which together with monthly reports from tire manufacturers and mass distributors gave fairly complete information concerning stocks of tires available for civilian consumption. In the interest of speed, accuracy, guaranteed reliability, and economy, the quarterly inventories were eventually taken by sampling. These advantages of sampling have given rise to increased usage and to the development of new theory,[1] but it should be stated that the sampling plan described here depends only on comparatively simple theory.

Purpose of this paper. The immediate aim of this paper is to describe the sample of dealers used in estimating the March 1945 tire-inventory; more specifically,

i. To state the reliability desired;
ii. To describe how this reliability was achieved through the application of basic principles and simple procedures;
iii. To describe how available knowledge concerning dealers' stocks was used to minimize the number of dealers in the sample and the work-load of weighting the results;
iv. To describe the estimation of the precision actually attained, and to compare the estimate of the actual precision with the aimed-at precision. This was done by examining a subsample of the returns;
v. To estimate the gains in efficiency accomplished by stratification.

The methods herein described may be profitably used in administrative problems of business and government relating to inventories, sales, employment, or traffic, wherever conditions are similar—i.e.,

* Now in the U. S. Office of Education.

[1] Morris H. Hansen and William N. Hurwitz, "Relative efficiencies of various sampling units," *Journal of the American Statistical Association*, vol. 37, 1942: pp. 89–94. "On the theory of sampling from finite populations," *Annals of Mathematical Statistics*, vol. xiv, 1943: pp. 333–52.

where a complete list of possible respondents is available along with other information such as size or type, which may serve as useful criteria for stratification.[2]

Definition of the universe. The universe to be covered was by definition the list of dealers on record with the OPA. This list contained the original registrations at the beginning of rationing (October 1942), plus dealers authorized subsequently ("new authorizations"), minus deletions that were made when notice was received that a business was defunct. Changes in name and address, and unreported defunct businesses, created the usual problems in interpreting nonresponses.

Evils of nonresponse; advantages of sampling. Nonresponses would be harmless if it were certain that a nonresponding dealer is out of business or has no tires. But unfortunately a nonresponse may mean other things as well—moved, change in name, or business as usual at the same address, coupled with inadvertance or inability to fill out the questionnaire. Every effort should therefore be made to keep the list trimmed of dead wood and to evaluate nonresponses by personal calls on a subsample.[3]

The minimum in nonresponses can be reached much easier with a sample than with a complete coverage because there are fewer of them, and the district offices are not bogged down with more calls on dealers than is humanly possible to make; and moreover, the sample almost always, by the principle of optimum allocation (vide infra), contains a preponderance of big dealers employing bookkeepers able to supply the figures and even taking pride in doing so. In illustration, the "complete count" of September 1944 resulted in 24,015 nonresponses or 17 per cent of the 140,989 questionnaires mailed out, whereas in the December 1944 and March 1945 samples the nonresponses were only 1.2 per cent and 3.9 per cent respectively of the 16,000 questionnaires mailed out, and be it noted that both these samples included a sample

[2] Information on size and type is not necessary for sampling, but ordinarily such information will make possible a considerable reduction in the size of sample required to attain the reliability desired. Stratification by area alone can be expected to show some small gains in sampling efficiency. Separation of dealers into groups widely different in type and sizes as of the last inventory will bring further gains, provided the inventories of dealers on the date of the sample still differ widely in type and size, in which case the additional efficiency gained by applying the principle of optimum allocation (vide infra) of the sample to the various strata is often striking. Fortunately, any large list must be arranged in some systematic fashion, such as by city, type, size, and order of receipt, whatever it is to be used for; and little or no additional labor may be required to group the dealers ideally for sampling.

[3] A systematic procedure for discovering the size of subsample of nonresponses to be followed up by a field interview has been worked out by William N. Hurwitz of the Census. This procedure elicits the greatest possible amount of information for a given allotted cost (mailed questionnaire plus field follow-up). It is being introduced into government surveys at considerable saving and gain in reliability. It was first expounded by Mr. Hurwitz at one of the Seminars in Statistical Inference at the Graduate School in Washington in 1943, and later published as Appendix C in the "Working plan for the annual census of lumber produced in 1943," which is obtainable from either the Census or the Forest Service.

of the nonresponses of September. Incidentally, the average number of tires per dealer was 50 per cent higher in both December and March for this group than the average inventory of all other dealers, thus pretty well deflating the possible interpretation in which refuge is so often taken, that nonresponses can be ignored on the pretext that they are average, or are composed mostly of small dealers, or represent no stock at all. These results bear out an interesting observation that had been made in tabulating the returns from the September complete count, viz., that the last 3000 dealers that reported subsequent to the second follow-up letter sent to delinquents, actually had on hand an average of over 10 tires per dealer against an over-all average of 6 tires per dealer for the 113,000 dealers that had already responded.

Some history of the quarterly inventories. The first inventory of tires was accomplished simultaneously with registration, following which quarterly inventories were taken, at first by the "cutoff" method by which small dealers were not asked to respond. Because the list of dealers showed not only names and addresses, but the number of tires by size and type on hand and consigned as of the latest complete inventory, it was possible to cut off any class of dealer, or to classify the dealers for efficient sample design. In June 1943 a complete count was attempted, following two inventories by cutoff that had been taken in December 1942 and March 1943, subsequent to the first registration (October 1942). The quarterly cutoff was repeated in September and December of 1943, but samples were taken in March and June 1944, followed by the attempted complete count of September 1944. At this time, the number of nonresponses was huge (24,015), and proper follow-up was impossible because of the sheer enormity of the task. Worrisome discrepancies appeared between the complete count of September 1944 and the earlier samples taken in March and June, the explanation of which turned out to be the more energetic and successful follow-up of the nonresponses in the samples. This explanation was not at first accepted, and indeed was not demonstrated until the sample of December 1944 was taken. Meanwhile, a study of these discrepancies had led to a re-examination of the entire reporting system, as a result of which sampling was recommended again for the December 1944 and March 1945 inventories.

II. PRELIMINARY CONSIDERATIONS AND THE STRATIFICATION OF DEALERS

Scope of this description. This article will be limited to a description of the sampling plan as revised for March 1945. The sample for the

previous quarter (December 1944) was planned along the same lines and will therefore be mentioned only as its results were used in the plans for March.

It is important to bear in mind that data were needed concerning several types and sizes of tires (of which there are new passenger tires, new truck and bus tires, motorcycle tires, used passenger tires and tractor-implement tires). Now a dealer in the sample was expected to report his stock of all types of tires, but owing to the differing dispersions of the various types of tires amongst dealers, some handling (e.g.) motorcycle tires, and some not, some handling truck and bus tires and some not, and some handling all types, a sample of dealers that will provide adequate precision for one type of tire might not do so for another. Analysis of the dispersions of tire-stocks demonstrated that a sample producing adequate results for new passenger tires would probably suffice for other types of tires with a few minor adjustments (Part IV).

The particular kind of sampling to be described here has these characteristics:

i. Stratified sampling;
ii. Optimum allocation of sample, to provide the aimed-at precision with the minimum number of questionnaires;
iii. Simple systematic selection of dealers within strata;
iv. A definite precision-requirement, consistent with the uses that were to be made of the data;
v. A predetermined procedure for weighting the results to get unbiased estimates;
vi. A determination of the precision that was actually attained, based on a sample of the returns (Part V).

The stratification. Study of past quarterly inventories revealed two main groups of tire dealers: Group A, containing dealers reporting stocks of new truck and bus tires in September 1944; and Group B, containing dealers reporting no new truck and bus tires but reporting stocks of new passenger tires in September 1944. This division was very effective in reducing the variance in the estimated total of new truck and bus tires. Similarly, it was effective for new passenger tires, because the larger dealers were automatically thrown into one group (Group A). This is so because dealers that handle new truck and bus tires usually aim to keep good stocks of all types of tires; actually, these dealers held about 85 per cent of the new passenger tires, as well as practically all of the new truck and bus tires. On the other hand, the dealers in Group B consisted mainly of small stores and service stations

handling relatively few popular size passenger tires usually incidental to the main business.

Besides Groups A and B there were five special groups, described in Part IV. Samples were drawn from these special groups, but they were relatively unimportant; the main consideration was Groups A and B, which contained practically all of the tires.

Dealers in each of these two groups were classified according to the numbers of tires that they held in September 1944 in intervals of 10 tires, i.e., 1–9 tires, 10–19 tires, 20–29 tires, and so forth. For Group A this classification was based on new truck and bus tires, and for Group B, on new passenger tires.

Method of expansion. From Class i, consisting of N_i dealers, a sample of n_i dealers was selected (cf. Remark 3). N_i was known; n_i was determined from considerations of optimum allocation as discussed further on. If S denotes the total number of (e.g.) new passenger tires on hand in March, then

$$S = N_1\mu_1 + N_2\mu_2 + \text{similar terms for the other subclasses through all groups} \tag{1}$$

where μ_i is the average number of new passenger tires held per dealer in Class i, which consists of N_i dealers. The sample gives, not μ_i, but $\bar{x}_i$ as an unbiased estimate thereof as is indicated in Remark 1. The comparison is disclosed by their definitions,

$$\mu_i = \frac{1}{N_i}(x_{i1} + x_{i2} + \text{similar terms through all } N_i \text{ dealers of Class } i) \tag{2}$$

$$\bar{x}_i = \frac{1}{n_i}(x_{i1} + x_{i2} + \text{similar terms through all the } n_i \text{ sample dealers of Class } i) \tag{3}$$

x_{ij} being the number of new passenger tires held by the j-th dealer in the i-th class. Because $\bar{x}_i$ is an unbiased estimate of μ_i it follows that

$$\hat{S} = N_1\bar{x}_1 + N_2\bar{x}_2 + \text{similar terms for the dealers of the sample} \tag{4}$$

is an unbiased estimate of S. The adjective "unbiased" signifies that if all possible samples of dealers were drawn by the method described, and if the responses were independent of the method of drawing the sample (which is not true in practice, except when samples are drawn from files of cards already containing the responses, or in an ideal experiment wherein the response is a number written on a chip), then the average value of all the estimates so obtained would be identical

with the results of a complete count. A mathematical definition of this concept is given below. The practical problem in designing a sample is to attain the desired band of sampling error for $\hat{S}$ at the lowest possible cost to business and government.

Remark 1. An unbiased estimate is one whose "expected" value is the quantity being estimated. Mathematically, an "expected" value is defined as the sum of the products of all values which could possibly occur in a given sampling plan multiplied by their respective probabilities of occurrence. Thus for the estimate $\hat{S}$ in Eq. 4,

$$\begin{aligned} E\hat{S} &= E(N_1\bar{x}_1 + N_2\bar{x}_2 + \text{similar terms through all classes}) \\ &= N_1E\bar{x}_1 + N_2E\bar{x}_2 + \text{similar terms.} \end{aligned} \tag{5}$$

Now from Eq. 3 it follows that

$$E\bar{x}_i = E\frac{1}{n_i}(x_{i1} + x_{i1} + \text{similar terms through all } n_i \text{ sample dealers in Class } i). \tag{6}$$

The probability of including any particular dealer in the sample is n_i/N_i, wherefore the sum of all possible sample values of inventories multiplied by their probabilities of occurrence is

$$\begin{aligned} E\bar{x}_i &= \frac{1}{n_i}\left(\frac{n_i}{N_i}\cdot x_{i1} + \frac{n_i}{N_i}\cdot x_{i1} + \begin{matrix}\text{similar terms through all}\\ N_i \text{ dealers in class } i\end{matrix}\right) \\ &= \frac{1}{N_i}(x_{i1} + x_{i1} + \text{similar terms through all } N_i \text{ dealers in Class } i) \\ &= \mu_i. \end{aligned} \tag{7}$$

Substitution of μ_i for $E\bar{x}_i$ in Eq. 5 gives Eq. 1. Thus the "expected" value of the estimate $\hat{S}$ is equal to S, which is the sum of the inventories of passenger tires of all dealers.

III. THE SAMPLE IN GROUPS A AND B

The principle of optimum allocation. There are two fundamental principles for guidance:

$$\frac{n_i}{N_i} = \frac{\sigma_i}{k} \tag{8}$$

$$\sigma_{\bar{x}_i}{}^2 = \frac{N_i - n_i}{N_i - 1}\,\frac{\sigma_i{}^2}{n_i}. \tag{9}$$

The first is the principle of optimum allocation,[4] which says that to obtain the greatest efficiency in sampling for (e.g.) passenger tires, the

[4] Optimum allocation was expounded by Neyman for simple stratified sampling in his article entitled, "On the two different aspects of the representative method," Journal of the Royal Statistical Society, vol. xcvii, 1934: pp. 558–625; pp. 570–80 in particular. Actually, Eq. 8 as given above is not strictly optimum allocation; the factor $\sqrt{N_i/(N_i-1)}$ on the right-hand side as given by Neyman has been omitted for convenience.

proportion (n_i/N_i) of dealers to be drawn into the sample from the i-th class should be proportional to the standard deviation σ_i of the passenger tires held by the dealers of this class on the date of the sample. The word "class" is used here in the general sense. It might mean a size-class within a group, or it might mean a whole group. The symbol k in Eq. 8 is a factor of proportionality and will be determined later by conditions fixed by the precision that is aimed at (see Eqs. 19 and 21).

Remark 2. Sometimes the total sample size $(n = \Sigma n_i)$ is fixed in advance by considerations of cost (and not, as here, by an aimed-at precision). For such problems it is convenient to know that the symbol k in Eq. 8 has the value

$$k = \frac{\sum N_i \sigma_i}{n} \tag{10}$$

in which case Eq. 8 gives

$$\frac{n_i}{n} = \frac{N_i \sigma_i}{\sum N_i \sigma_i} \tag{11}$$

This equation fixes the sample n_i within a class in terms of the total sample n, supposed known. In the problem of this article, the approach was from the other direction—a desired precision was aimed at, and k determined accordingly (cf. Eqs. 19 and 21). The sample sizes n_i were then computed from Eq. 22. They of course satisfy Eq. 11, but n was not known until the n_i were computed and added up.

In symbols, the standard deviation σ_i in Eqs. 8 and 9 is defined by the equation

$$\sigma_i^2 = \frac{1}{N_i} \sum_{j=1}^{N_i} (x_{ij} - \mu_i)^2. \tag{12}$$

Eqs. 8 and 9 are intended to refer to conditions existing at the date of inventory. Actually, σ_i is known only for some earlier date (in this case, the preceding September and December), so it is necessary to approximate σ_i for the date of the sample (cf. Eq. 14). Eq. 9 gives the variance of the mean number of tires per dealer as would be obtained in an ideal experiment in which repeated samples of n_i dealers are drawn at random and without replacement from the N_i dealers. Each class of dealer will contribute a term (see Eq. 13 ahead).

Remark 3. The n_i dealers from any class were selected systematically by starting at a random point in the first interval in the file and taking out one card at a time at a constant interval throughout the file for that class. A systematic selection is much easier to carry out than any serious attempt at random selection. Moreover, systematic selection in this type of problem may introduce slight gains over a random selection, for the reason that systematic selection assures nearly proportionate representation of dealers

from each city and state, because of the arrangement of the files. This may have been of some importance because the inventories in one broad locality usually differ from those in another both qualitatively and quantitatively. On the other hand it must also be recognized that systematic selection will sometimes introduce losses. The theoretical calculations made by the Madows have not been applied to see exactly what the gains and losses of systematic sampling might be in this type of problem. The equations for random sampling are used instead under the assumption that they indicate the approximate level of the variance of this design. It should be noted that even if the variance of the systematic design were as much as 25 per cent higher than the variance of a random sample of equal size, the usefulness of the inventory obtained by the methods of this paper would hardly be affected. The authors have had the benefit of guidance from Dr. Madow in these questions.

An assumption for arriving at useable values of σ_i. The symbol $\hat{S}$ in Eq. 4 denotes the total number of new passenger tires as estimated from a sample. By the law of propagation of variances[5] the variance of $\hat{S}$ will be

$$\begin{aligned}\sigma_S^2 &= N_1^2\sigma_{\bar{x}_1}^2 + N_2^2\sigma_{\bar{x}_2}^2 + \text{similar terms for all other classes} \\ &\qquad\qquad \text{through all groups} \\ &= N_1^2\,\frac{N_1 - n_1}{N_1 - 1}\,\frac{\sigma_1^2}{n_1} + N_2^2\,\frac{N_2 - n_2}{N_2 - 1}\,\frac{\sigma_2^2}{n_2} + \text{similar terms} \\ &\doteq N_1^2\left(1 - \frac{n_1}{N_1}\right)\frac{\sigma_1^2}{n_1} + N_2^2\left(1 - \frac{n_2}{N_2}\right)\frac{\sigma_2^2}{n_2} \\ &\qquad + \textit{similar terms.}\end{aligned} \tag{13}$$

The second line follows from the first by the introduction of Eq. 9, and the third line is simply a close approximation obtained by dropping the -1 in each denominator. Later on, Eq. 13 will be used as it stands for the evaluation of the precision actually attained; this will be done by evaluating each σ_i from a subsample of the returns (Part V). Meanwhile, though, in the planning of the sample, suitable values of σ_1, σ_2, etc. must be settled upon from considerations of possible trends and dispersions of tire inventories. For the March sample there was the experience of the December sample to fall back on, which helped considerably.

It seems reasonable to suppose that the changes in dispersion taking place in the inventories of a group of dealers over a 3 or 6 months' period will be roughly proportional to their average inventory of the

[5] Propagation of variance is treated in many books on least squares and curve fitting; for example, Ch. III in Deming's *Statistical Adjustment of Data* (Wiley, 1943).

base date (Sept. '44). This is only saying that the stock of a large dealer will usually change more in either direction than the stock of a small dealer, and about in proportion to size as measured by initial inventory. Indeed, Messrs. Hansen and Hurwitz, in empirical tests on many lines of work, have found this to be a very useful assumption. It of course breaks down in the small size-classes (cf. the treatment of the smallest subclass in Table IV; and the treatment of "O-blocks" in the Part VI). With this in mind, it was perceived that in the returns from the December sample, σ_t/μ_t was roughly constant in the different size-classes of Group A and hardly exceeded 1.25. Similarly, it hardly exceeded .75 in Group B. To allow for still further dispersion during the additional 3 months between December and March it was assumed that

$$\left.\begin{aligned} \frac{\sigma_t}{\mu_t} &= 2 \text{ in Group A} \\ \frac{\sigma_t}{\mu_t} &= 1 \text{ in Group B} \end{aligned}\right\}. \tag{14}$$

These approximations are admittedly rough, as is only too obvious from Tables I and II.

TABLE I

COEFFICIENTS OF VARIATION FOR SEPTEMBER AND DECEMBER FOR GROUP A

GROUP A: DEALERS THAT IN SEPTEMBER HAD NEW TRUCK AND BUS TIRES, BUT FEWER THAN 40 {NEW PASSENGER TIRES, USED TRUCK AND BUS TIRES, USED PASSENGER TIRES

μ_t and σ_t were estimated from a subsample of about 100 returns taken systematically from each size-class

No. of new truck and bus tires in Sept.	NEW TRUCK AND BUS						NEW PASSENGER					
	September			December			September			December		
	μ	σ	σ/μ	μ	σ	σ/μ	μ	σ	σ/μ	μ	σ	σ/μ
1– 9	3.5	2.3	.66	4.1	5.9	1.44	8.9	9.3	1.04	[illegible]	18.2	[illegible]
10–19	13.5	2.7	.20	13.0	9.6	.74	11.1	10.5	.95	21.0	26.3	1.25
20–29	24.1	2.9	.12	25.0	13.2	.53	13.6	11.2	.82	34.2	40.6	1.19
30–39	34.0	2.5	.07	38.2	17.9	.47	18.5	12.1	.65	34.2	28.2	.82

Remarks regarding assumptions. Three remarks are appropriate here. The most efficient use (optimum allocation) of a given number of questionnaires only requires that the *ratios* $\sigma_1:\sigma_2$, etc. between the standard deviations σ_t be assigned, and not their absolute magnitudes. Now very often one can make good guesses at these ratios even though the value assigned to each σ_t is too high or too low. Of course, if one consistently

assigns too high a value to every σ_i, and computes the size of sample required to give the precision aimed at, the precision actually attained will be needlessly high, which is not desirable as it entails a heavier burden of response and heavier cost than was actually necessary. On the other hand, if one consistently assigns too low a value to each σ_i, the precision actually attained will not be as good as the precision aimed at, though it may nevertheless be good enough. In the second place, it is a fact that the efficiency of the sample (as measured by the inverse of σ_S^2) is only feebly altered by moderate departures from the optimum ratios between n_1, n_2, etc., as given by Eq. 8, wherefore a sampling plan may still be very efficient even if some of the assumptions and approximations are crude.

In the third place, regardless of what assumptions and approximations were made, or why, there need be no guess-work in the final results, because an analysis of the returns will show what precision was actually attained (cf. Part V).

TABLE II

COEFFICIENTS OF VARIATION FOR SEPTEMBER AND DECEMBER FOR GROUP B

GROUP B: DEALERS WITHOUT NEW TRUCK AND BUS TIRES IN SEPTEMBER BUT WITH

FEWER THAN 40 {USED PASSENGER TIRES / USED TRUCK AND BUS TIRES}

μ_i and σ_i were estimated from a subsample of about 100 returns taken systematically from each size-class

No. of new passenger tires in September	NEW TRUCK AND BUS						NEW PASSENGER					
	September			*December*			*September*			*December*		
	μ	σ	σ/μ	μ	σ	σ/μ	μ	σ	σ/μ	μ	σ	σ/μ
0	—	—	—	.2	.9	4.5	—	—	—	1.0	3.6	3.60
1–9	—	—	—	.5	1.8	3.6	4.0	2.4	.60	6.7	8.2	1.22
10–19	—	—	—	1.0	3.3	3.3	13.1	3.1	.24	13.0	9.9	.76
20–29	—	—	—	1.2	4.9	4.1	24.0	2.7	.11	24.7	11.4	.46
30–39	—	—	—	.8	2.7	3.4	33.9	2.0	.06	32.0	12.4	.39

Evaluation of k. To find k we note that Eq. 8 gives

$$\frac{\sigma_i^2}{n_i} = \frac{k\sigma_i}{N_i} \tag{15}$$

which when used in Eq. 13 shows that

$$\sigma_S^2 = k\left\{\frac{\sigma_1}{\mu_1} N_1\mu_1\left(1-\frac{n_1}{N_1}\right) + \frac{\sigma_2}{\mu_2} N_2\mu_2\left(1-\frac{n_2}{N_2}\right) + \text{similar terms}\right\} \tag{16}$$

Now suppose that we restrict consideration to Groups A and B, fixing the aimed-at precision so that it should be attained from these two groups alone. The relatively small contributions of passenger tires from other groups of dealers (Part IV) will only slightly increase the precision. This procedure is not strictly optimum allocation over all strata, but is not far from it. One further simplification will be made; the finite multiplier $1-n_i/N_i$ in each term of Eq. 16 will be dropped (i.e., assumed to be near unity, as it would be if the sampling ratio n_i/N_i were small). Actually, the calculations for k (vide infra) should perhaps have been repeated to get a second approximation by retaining the finite multipliers in Eq. 16 with rough values of n_i/N_i. The coefficients of S_A and S_B in Eq. 17 would thereby have been decreased, but it was supposed that the effect was slight and not worth the effort. The next step, therefore, was to ignore the finite multipliers and factor out σ_i/μ_i for Groups A and B separately in Eq. 16 to find that

$$\begin{aligned} \sigma_S{}^2 = k\Big\{&\left(\frac{\sigma}{\mu}\right)_A (N_1\mu_1 + N_2\mu_2 + \text{similar terms in Group A}) \\ &+ \left(\frac{\sigma}{\mu}\right)_B (N_1\mu_1 + N_2\mu_2 + \text{similar terms in Group B})\Big\} \\ = k\{2S_A &+ S_B\} \end{aligned} \tag{17}$$

because $\sigma_i/\mu_i = (\sigma/\mu)_A = 2$ for Group A and $\sigma_i/\mu_i = (\sigma/\mu)_B = 1$ for Group B according to a previous decision (Eq. 14). It follows that the

$$(\text{C.V. } \hat{S})^2 = \frac{\sigma_S{}^2}{S^2} = k\,\frac{2S_A + S_B}{S^2} \tag{18}$$

where C.V. $\hat{S}$ denotes the coefficient of variation of $\hat{S}$, or its standard error expressed in units of S itself. Solved for k this equation gives

$$k = \frac{S^2(\text{C.V. } \hat{S})^2}{2S_A + S_B}. \tag{19}$$

It is now necessary to anticipate the division of the total number of tires S between Groups A and B. For simplicity it was assumed that the division in March would be what it was in December, viz.,

$$S_A : S_B = 8:1. \tag{20}$$

Moreover, it was necessary to decide on a suitable value for S as of March '45, and after some deliberation it was decided that 1.8 million could not be far wrong. This number fixed the values of S_A and S_B as

1.6 and .2 million respectively, whereupon k could be obtained from Eq. 19 as soon as the aimed-at precision C.V. S was decided upon (next section).

TABLE III

SIZE OF SAMPLE FOR GROUP A

(Group A is defined in Table I)

No. of new truck and bus tires in Sept.	*p_i For new passenger tires; assumed for March*	*Size of sample computed (Eq. 28)*	*Size of sample decided upon*	*Number of dealers in the Universe N_i*	*Number of dealers in the Sample n_i*
1–9	15	15%	1 in 6	10,850	3,317
10–19	22	20	1 in 5	3,250	650
20–29	30	30	1 in 3	1,613	538
30–39	35	35			
40–49	45	45	1 in 2	894	447
50–59	55	55			
60 and over	—	100	all	1,662	1,662
			Total	27,209	6,275

TABLE IV

SIZE OF SAMPLE FOR GROUP B

(Group B is defined in Table II)

Number of new passenger tires in September	*p_i For new passenger tires; assumed for March*	*Size of sample computed (Eq. 28)*	*Actual size of sample decided upon*	*Number of dealers in the Universe N_i*	*Number of dealers in the Sample n_i*
0	1	*½%	1 in 20	6,340	317
1–9	7	3½	1 in 25	23,050	940
10–19	13	6½	1 in 12	6,500	542
20–29	25	12½	1 in 8	1,893	220
30–39	35	17½	1 in 5	1,130	227
40–49	45	22½			
50–69	60	30	1 in 3	358	119
70–99	65	42½	1 in 2	163	82
100 and over	—	100	all	108	108
			Total	40,148	2,630

* The sample decided upon was larger than computed for this size-class not only because its dispersion was much higher than the other size-classes (see Table II) but because this class of dealer is subject to more important relative shifts over time.

The aimed-at precision. It was decided that a coefficient of variation of 1.5 per cent would be desirable. This corresponds to a 3-sigma error-band of 4 or 4½ per cent outside of which errors of sampling practically never fall. This band of error seems to be small enough in view of the uses to be made of the data, and in view of other errors and biases that arise from difficulties in counting tires on hand and consigned.[6] Also

[6] An attempt at classifying a number of errors and biases was published by W. Edwards Deming in a paper entitled "On errors in surveys," *American Sociological Review*, vol. ix, 1944: pp. 359–69.

there was the stock of unknown magnitude held by unauthorized dealers. Altogether, further refinement of sampling error seemed out of place, because it would entail additional burden of response on the part of dealers, and increased costs to the government, and would moreover enhance liability to errors arising from nonresponse and other sources, most of which, as surveys are actually carried out, become more troublesome as the sample increases, and reach their maximum in attempts at complete counts.

With the aimed-at coefficient of variation of $\hat{S}$ set at 1.5 per cent, Eq. 19 gave

$$k = \frac{1.8^2 \times 10^{12} \times 1.5^2 \times 10^{-4}}{2 \times 1.6 \times 10^6 + .2 \times 10^6} = 214 \tag{21}$$

If the finite multipliers in Eq. 16 had not been neglected, k would have turned out to be a little greater than 214. Nevertheless, for ease in computation k was fixed at 200, whereupon from Eqs. 8 and 14

$$\left.\begin{aligned} \frac{n_i}{N_i} &= \mu_i \text{ directly in per cent for Group A} \\ &= \tfrac{1}{2}\mu_i \text{ directly in per cent for Group B} \end{aligned}\right\} \tag{22}$$

And thus the sample was allocated in Groups A and B, which account for most of the new passenger tires (truck and bus also). Tables III and IV show the calculations and recommended sampling ratios.

IV. SPECIAL GROUPS OF DEALERS

The nonresponses of September. In September there were 24,915 questionnaires not returned, even after two "follow-up" notices. In December a 4 per cent sample of these names was drawn: in number this sample was 997. Here is what happened.

Number of questionnaires mailed out as a 4% sample of the September Nonresponses			997
Failed to return by mail and personal visit, or could not be located, or no personal follow-up made			310
Returned			687
Marked out of business	217		
Marked 0 passenger tires, or completely blank (interpreted as having no stock)	134	470	
Reports showing 1 or more new passenger tires	336		

For March, a 5 per cent sample of the September nonresponses was recommended. The coefficient of variation (37.9/11.9 or nearly 3) of this group perhaps called for a slightly larger sample but it was deemed better not to risk overloading the district offices with this group (which was obviously inclined not to respond) and to accept a little larger sampling error at the gain of decreased bias of nonresponse.

For the 470 dealers returning blanks or 0 or 1 or more tires, the mean number of new passenger tires per dealer was 11.9, with a standard deviation of 37.9. When expanded by 25 this group gave

$$11.9 \times 470 \times 25 = 140{,}000 \text{ new passenger tires.}$$

This figure is to be interpreted as a minimum in the number of passenger tires held by the nonresponses of September because the 310 that failed to return likely also held tires. Anyhow, this minimum of 140,000 was a twelfth of all the new passenger tires in the hands of dealers in December and if these dealers held anything like the same stocks in September it is easy to see why there should have been discrepancies between the September complete count and the two quarterly samples in the preceding March and June (as was mentioned earlier). The ability of the staff of the district offices to call on the smaller number of delinquent dealers in a sample is responsible for the better response in December and March.

Blanks. In September, 29,133 dealers returned their questionnaires blank. A sample of 3 per cent of these was used in December, which when expanded on the basis of the returns accounted for 71,000 new passenger tires. One might conjecture that this group of dealers held enough tires in September to aggravate the trouble of balancing the books. A 3 per cent sample was recommended again for March.

New authorizations. There were 1,000 new authorizations between September and December and another 1,000 were allowed for over the interval between December and March. A sample of 200 was recommended as being adequate and easy to draw, being 10 per cent.

Manufacturers' outlets. There are about 2,000 manufacturers' outlets, averaging for September about 75 new passenger tires per dealer. A 100 per cent sample was recommended but a 1 in 2 or 1 in 4 sample might have been good enough because they were afterward observed to have fairly uniform stocks. No calculation of their coefficient of variation (σ/μ) was made.

Dealers holding in September 40 or more used tires of any kind but fewer than 40 new passenger or new truck or bus tires. A 25 per cent sample

was recommended: every 4th card starting with the 2d, first classifying the dealers in intervals of 10 tires (40–49, 50–59, etc.).

Motorcycle tires. Of the dealers reporting no new passenger tires and no new truck and bus tires in September, and fewer than 40 used passenger or truck and bus tires, 296 reported new motorcycle tires, and 91 reported used motorcycle tires. After inspecting the reports of these 387 dealers the sampling plan shown below was recommended:

No. of motorcycle tires in September	*Sample*
1–4	25% or 1 in 4
5–9	25% " "
10–14	25% " "
15–19	25% " "
20 and over	100% " all

Of course there were motorcycle tires held by other dealers, but stocks of motorcycle tires were well dispersed and the precision in this category should be good enough.

The contributions to the variance from these special groups was small because they were sampled adequately (approximately according to optimum allocation) and also because Groups A and B contained nine-tenths of the new passenger tires (and even more of the new truck and bus tires). In particular, the manufacturers' outlets, like the dealers with heavy stocks in Groups A and B, contributed no variance, being sampled 100 per cent.

Size of sample—all groups. The over-all sample from all groups, as calculated, turned out to be 14,750 dealers, which would have been about 1 dealer in 10. However, partly by mistaken carry-over of instructions from previous samples, and partly to gain simplicity in operation, bigger samples than specified were taken out of some of the special classes and the actual number of questionnaires mailed out was close to 16,000. As mentioned earlier, the response from the March sample (with two follow-up letters) was 96.1 per cent, and this sample included a 5 per cent sample of the dealers that did not respond in September.

V. RESULTS. COMPARISON OF THE PRECISION ATTAINED WITH THE PRECISION AIMED AT

The precision attained. The effect of stratification. After the returns were in, a subsample of about 100 cases was drawn systematically from every group for the purpose of estimating σ_i and μ_i for use in Eq. 13 whence the precision of the results was obtained. For comparison, the coefficients of variation were also computed for the precision as it would

have been if no stratification had been introduced. In both investigations the assumption was made that the equations of random sampling can be applied to estimate the precisions of our systematic samples, as in Remark 3. The apparent coincidence between the precision aimed at for the new passenger tires (1.5 per cent) and the precision actually attained (again 1.5 per cent), is partly an illusion, because the precision aimed at applies strictly only to Groups A and B, whereas the precision actually attained and shown in the table was computed for all groups combined. However, as was explained earlier, it was known that the contributions from the other groups could not have much effect on the over-all precision, so it can be concluded from the table that the agreement was excellent—unfortunately too good for illustrative purposes. The outcome does nevertheless show qualitatively the kind of results that will be obtained from plans that are drawn up with reasonable care, even though some of the assumptions and simplifications that one is forced to make often appear crude at the time. In this connection it should be borne in mind, too, that in practice it is never important to hit the aimed-at precision very closely, but it is important to learn from the returns what the precision actually was. This can usually be done at a small investment, and should never be omitted. A further advantage of studying the returns is increased knowledge and experience to be applied in future samples.

The precision for the new truck and bus tires (1.4 per cent; see Table V) turned out to be entirely satisfactory, as was expected, even though the criterion for fixing k and hence for fixing the size of the sample was based on new passenger tires. The reason is that the truck and bus tires were heavily concentrated in Group A.

TABLE V

ACTUAL COEFFICIENTS OF VARIATION ATTAINED OVER ALL DEALERS

Plan	*New truck and bus tires*	*New passenger tires*
Stratified with optimum allocation	1.4 per cent	1.5 per cent
Unstratified (for comparison)	4.4	2.5

Interpretation of sampling error. The observed coefficient of variation, measured after the returns are in, tells us that the number of new passenger tires held by dealers on March 31st, as determined by the sample, can hardly differ from the results of a "complete count" by more than $4\frac{1}{2}$ per cent (three coefficients of variation). The error band for truck and bus tires happens to be about the same width or smaller. But to

interpret these error bands it must be understood that the theoretical complete count with which the sample is compared must be taken with the same thoroughness as the sample, including pressure on the non-responses. Moreover, the complete count must cover the same universe of dealers, and not include (e.g.) unauthorized dealers. Thus, while the sample contains sampling errors, and is therefore less precise than a complete count, it is undoubtedly more accurate than a complete count would have been in practice.

VI. NOTE IN REGARD TO STRATIFICATION AND CONSERVATION OF INFORMATION FROM COMPLETE COUNTS

It is obvious from Table V that the stratification based on the complete count of September increased the efficiency of sampling enormously, even though considerable dispersion of inventories took place between September and December, and still more between December and March. Unstratified, the required sample for new passenger tires would have been 2½ times as large; and for the new truck and bus tires it would have been 5 times as large. It should be explained that the gains in efficiency attributed here to stratification were actually the gains of stratification with optimum allocation. The gains that would have been made by stratification with proportionate sampling would not have been as great—perhaps two-thirds as great in this survey.

This experience seems to be typical in many sample surveys and teaches an important lesson in regard to conserving whatever information is available concerning the universe. Thus, the 1940 Census of Population providing statistics by blocks, tracts, and city size groups, even though now out of date in many spots because of heavy population shifts is nevertheless still effective for purposes of stratification, even for sampling blocks within cities, as experience and theory demonstrate. Similar remarks hold for the 1939 Census of Business. Some of the reasons are found in the section entitled "Remarks regarding assumptions." For example, optimum allocation will indeed be obtained even though the census or latest available complete count only indicates relative sizes (of firms, blocks, farms, etc.) and is entirely wrong on absolute values. But this is not all; a complete count, although out of date, will often go a long way toward segregating areas that may have suffered important changes, and which by the principle of optimum allocation should be sampled in higher proportion. Thus, in the footnote to Table IV it is explained that the O-class was sampled heavier than was indicated by the rule that was used for the other classes. Similarly, in the sample censuses of certain congested cities

taken in the spring of 1944, the blocks were stratified into classes according to the number of dwelling units in 1940, and every n-th block taken out for the sample. But the "O-blocks" (those that had no dwelling units in 1940) were thrown into a separate class for special investigation as a safeguard against the enormous changes that can take place when a large housing development goes up on land not previously occupied. Another illustration of the conservation of available information is found in the best practice in sampling in industry, wherein the lot, shipment, shift, or even order of receipt, are usually effective for subgrouping even though it may appear that the product must have become thoroughly mixed following production. It never is thoroughly mixed unless there has been a deliberate and skillful attempt to mix it.

It is not to be inferred that promiscuous stratification will get results. After the main assignable causes of variability have been removed through stratification, further stratification may avail too little to repay the cost.

It is a pleasure for the authors to record their indebtedness to Messrs. Morris H. Hansen and William N. Hurwitz for their early work with the authors on the design of the sample as it was first used for the tire inventory. It is a pleasure also to record the splendid cooperation of Messrs. S. Schotland and M. E. Robbins who were in charge of the Inventory and Control Branch of the Office of Price Administration in New York, where the samples were drawn and tabulated. Valuable professional assistance was rendered throughout the work by Dorothy Gottfried and Miss Sylvia Wolosoff of the New York Branch.

WAR-TIME ALUMINUM STATISTICS*

MORTIMER D. [illegible]
Reconstruction Finance Corporation

THE START of the national defense program in 1940 found the United States government with few weapons and, at least in the case of aluminum,[1] little knowledge of the raw materials required to make those we needed. Moreover, the government had virtually no information to show the extent to which the United States possessed fabricating capacity to fulfill whatever the material requirements might be. Before the war, the Bureau of Mines collected annual data on virgin aluminum production and secondary recovery (from scrap), and the Department of Commerce tabulated our imports and exports, but that was about all. The lack of a comprehensive governmental statistical program for the aluminum industry coupled with the absence of a trade association program that might have served as a temporary substitute meant not only that new statistical surveys had to be conceived, put into operation, and refined, but also that appropriate research techniques and methods of analysis had to be worked out. The development of these techniques and methods under the uncompromising pressure of the war will be the major theme of this presentation.

The fact that in 1940 the Aluminum Company of America (Alcoa) was the only producer of virgin aluminum as well as the principal manufacturer of aluminum semifabricated products facilitated the task of obtaining the information needed to begin the business of national defense planning. Even with this central source of information the task was far from a painless one, and it was not until well into 1942 that lack of adequate aluminum statistics ceased to be a major source of difficulty in the work of the War Production Board (WPB). Although a comprehensive statistical reporting system was instituted by the

* The opinions expressed are those of the author and are not necessarily representative of the official views of the War Production Board, with which the author was associated, or of the Reconstruction Finance Corporation.

[1] *Aluminum* is used as a generic term to apply to all of the products of the aluminum industry, whether made of commercially pure aluminum or aluminum-base alloy and whether semifabricated or unfabricated. *Aluminum metal*, *aluminum pig*, and *aluminum ingot* are often used interchangeably to represent unfabricated metal. *Pig* is better used, however, to denote a simple form of unfabricated metal which, when refined and cast in a special form for subsequent use, becomes an *ingot*. Virgin metal (sometimes referred to as *primary*) is aluminum made from alumina, while scrap metal (commonly called *secondary*) is aluminum made from scrap. Some authorities use the term [illegible] to apply only to aluminum made from scrap whose original alloy identity has been lost, but this is a fairly restricted use of the term. Either *semi-fabricated products*, *aluminum products*, or *fabricated products* can be employed to represent the products of the industry which have been processed into desired alloy compositions, forms, and shapes; for example, sand castings, sheet, foil, extrusions, powder. In this paper *finished products* and *end products* will denote manufactures, such as aircraft and cooking utensils, made in whole or in part from aluminum semi-fabricated products. The United States could use a glossary of standardized aluminum terminology.

Office of Production Management, WPB's immediate predecessor, before Pearl Harbor, there was no substitute for background data and no telescoping the time for trends to make themselves known. Further, the aluminum industry proved to be far from a simple organism that revealed all of its structure and function under cursory examination. And it was found that the accounting and production planning departments of big business are not so flexible that they can readily conform to governmental reporting requirements that must be tailored to fit also the potentialities of some 1500 small aluminum establishments.

INDUSTRY STRUCTURE AND COMMODITY CLASSIFICATION

The activities of WPB in channeling the flow of materials, scheduling production, and guiding the operations of industry generally required, for administrative even if not for statistical considerations, the careful definition of the commodities and industries under the cognizance of each of the Board's separate organizational units. Because of the preponderance of the productive capacity encompassed by Alcoa and the Reynolds Metals Company and the close integration of operations within each company, it was natural to define the aluminum industry so as to make it coincide as far as possible with the boundaries of the activities of its two principal members. At the same time the production and accounting practices of the smaller producers had to be taken into account.

This approach proved on the whole to be satisfactory for both administrative and statistical purposes and brought the aluminum products and facilities shown in Figure 1 within the definition of the industry. Alcoa produced all the products included; Reynolds, most of them. A limited number of products made by Alcoa or Reynolds either were too widely manufactured and distributed as finished products or were too difficult to control as aluminum products to be included; for example, screw machine products, collapsible tubes, so-called Hooker tubing, processed foils (laminated, printed, embossed), and stampings. Empirical judgment would indicate that rivets, cable, and pigment-type powders, all of which resemble finished components as much as semi-fabricated products, might also have been excluded from the definition. Distributors of aluminum products and dealers in aluminum scrap completed the industry's coverage.[2]

[2] The mining and beneficiation of bauxite (the only mineral commercially used as a source of aluminum) and the production of alumina (aluminum oxide, extracted from bauxite and reduced to aluminum electrolytically) could be included as part of the industry under a general definition, but non-metallic materials such as these, so far removed from munitions production, were subjected to their own distribution and control techniques in the Board.

THE ALUMINUM INDUSTRY
STRUCTURE AND FLOW OF METAL
SIMPLIFIED

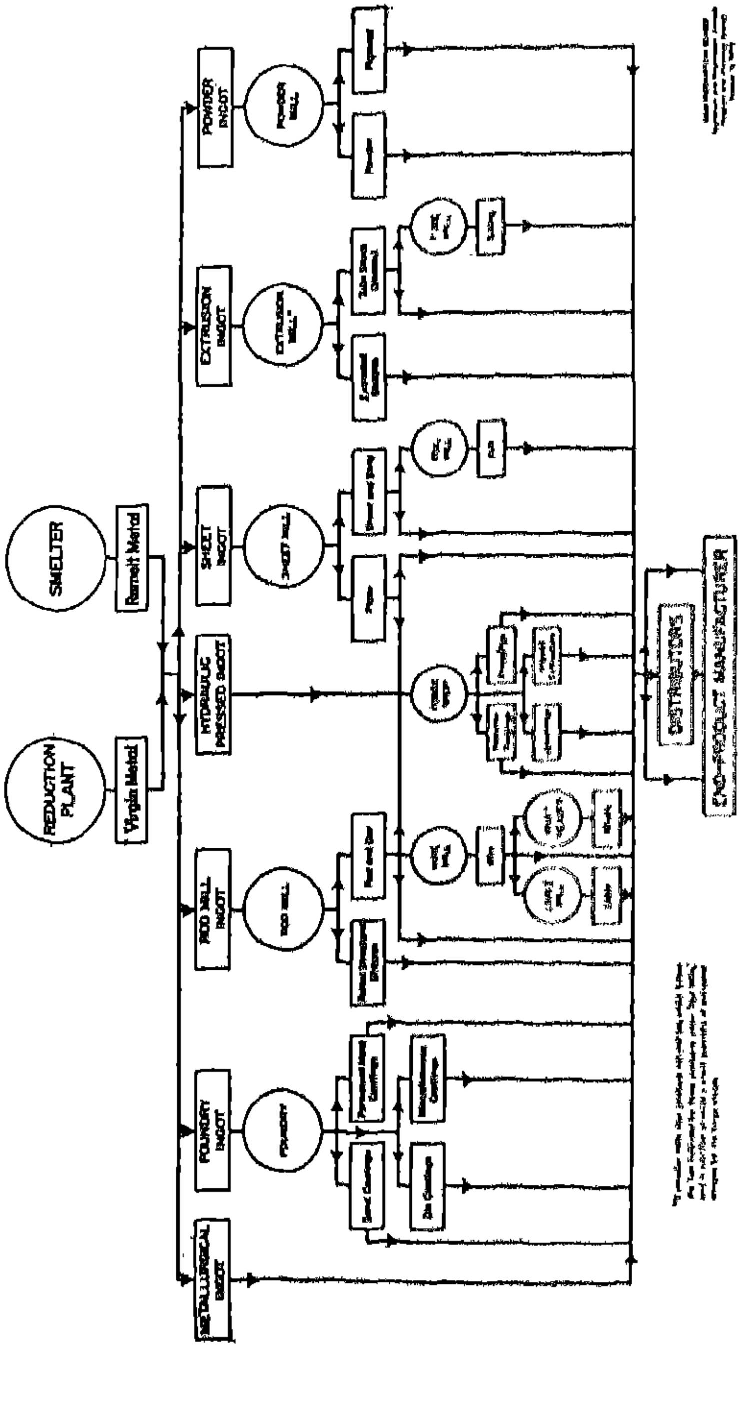


FIGURE 1

While Figure 1 indicates the principal headings under which aluminum products can be classified, it does not reveal the detail to which useful classification could be and was refined in the administration of the aluminum program. It is almost a truism to say that the greater the detail available on requirements and supplies the better they can be matched. For instance, early in the war rod of small diameter became critical, but without knowledge of capacity, production, and requirements according to diameter, it was difficult to plan for the relief of the shortage. Moreover, plant capacities were determined to a great extent by the alloys and other specifications called for by consumers. The capacity of an extrusion plant is much greater poundage-wise when working on heavy or soft alloy sections than when producing light or hard alloy sections; and the capacity of a foundry falls when specifications are so rigid that ordinary practices of repairing defects in castings are outlawed.

The decision as to the aluminum classification scheme to be used by WPB in the operation of the Controlled Materials Plan—the plan which governed the distribution of aluminum, steel, and copper from 1943—became a contest with the statisticians and supply planners on one side and the administrators of the overall plan on the other. The former wanted a detailed classification broken down by alloy, size, and heat treatment; the latter were fearful that a detailed classification would unduly burden industry with paper-work and cause the plan to become enmeshed in "red-tape" and break down. The result, as might be expected, was a compromise. Instead of over seventy categories of aluminum—the maximum number ever proposed—it was agreed that bills of materials (the ultimate source of requirements data) would be prepared by industry on the basis of a thirty-three-item classification, that requirements would be calculated and submitted by the procuring agencies on the basis of a more consolidated list to vary according to the agency, and that allotments would be issued in eight categories.[3] The eight categories were also used for the public release of aluminum shipment statistics by WPB in the *Facts for Industry* series.

The limitations imposed with respect to allotment control and requirements data did not extend to the collection of statistics on the operations of the aluminum industry. Accordingly, throughout the war, data on aluminum production, shipments, and past-due unfilled orders

[3] The eight categories used for aluminum allotments compare with two in steel and six in copper. As experience was gained, the aluminum classifications were condensed until requirements were presented in eight categories and allotments were issued in one. The aircraft industry, by far the principal consumer of aluminum, was subjected for some time to more detail than other consumers, however, both in terms of requirements calculation and allotment accounting.

were collected in more detail than the eight categories, for it was found that analysis of shipments in terms of alloys, sizes, and heat treatments revealed significant trends in the pattern of business[4] that were invaluable in laying out production and facility expansion programs. In a sense, detailed statistical reporting and analysis of the past became a substitute for detailed controls over the future. As aluminum became easier in supply a good deal of this detail was dispensed with and the reporting job of industry was reduced as much as possible. The number of reporting categories was gradually condensed from forty-two to thirty-one.[5]

REQUIREMENTS DEVELOPMENT AND ANALYSIS

The fulfillment of the requirements of the armed forces, the foreign nations entitled to Lend-Lease assistance, and the war-supporting domestic economy for materials and finished products constituted the core of the responsibility of WPB. The Board delegated the responsibility for the compilation of requirements to the so-called "claimant agencies"—the War and Navy Departments, the Foreign Economic Administration, the Office of Defense Transportation, the National Housing Administration, the Petroleum Administration for War, and the other Federal agencies under whose cognizance the economic potential of the country was mobilized. WPB assumed the representation of those industries which produced components of such wide usage that they could not be handled by any one of the other agencies. The WPB Office of Civilian Requirements represented the civilian population.

Under CMP, the calculation of requirements began with the placement of a prime contract for an end-product, bearing specifications in terms of materials, performance, or both. The prime contractor prepared a bill of materials showing the amount of copper, aluminum, and other commodities, in designated shapes and sizes, required to produce a unit of the end-product, making appropriate allowances for scrap losses, spoilage, and spare parts. (A bill of materials was also prepared in some circumstances by the subcontractor and was submitted to the prime contractor for incorporation in his bill of materials.) The summary bill of materials was transmitted by the prime contractor to his claimant agency, where it became one of the many used in calculating

[4] The phrase *pattern of business* is used to represent the composition of [illegible] according to the sizes, gauges, alloys, heat treatments, finishes, tolerances, and other [illegible]

[5] The aluminum section of the U. S. Standard Commodity Classification [illegible] from the scheme used by WPB, but the two can be brought together on a [illegible] basis.

the materials required for that agency's aggregate of procurement programs. The claimant agency compiled the material requirements for all its programs, taking into account the lead-time (length of manufacturing cycle) for each end-product and the expected change in contractors' unprocessed and in-process stocks, and submitted the resulting figures to WPB. The requirements were submitted quarterly and included not only the data for the one quarter under immediate consideration, but also for as many as five additional quarters.

The Board, with a staff of economists, statisticians, and industrial specialists, analyzed or "screened" the requirements submitted by each claimant to determine initially whether the requirements were reasonable representations, in terms of materials, of the agency's procurement programs. When total screened requirements exceeded available supplies, the Board had to investigate the conflicting demands, weigh the consequences of alternative program reductions on each of the armed forces, on the civilian economy, and on the Lend-Lease recipients, and decide how the available supplies should be allocated. On some occasions, when the competition for materials was finally narrowed to the armed services, the question as to the programs to be cut back was referred to the Joint Chiefs of Staff or the Under Secretary of War. The final decision of the Board was embodied in formal allotments of each commodity, issued by the Chairman of the WPB Requirements Committee to the various claimants.

This lengthy chain of events had a marked effect on bringing material requirements out of the chaos in which the beginning of the war found them. That is not to say, however, that the requirements were wholly accurate descriptions of the materials really needed. Figures on aluminum, for example, show that for nearly every quarter of the operation of CMP, shipments were substantially below originally stated requirements and allotments, and yet past-due unfilled orders showed no substantial change. It is evident that requirements were usually inflated. There are many reasons why this was so, and a few will be enumerated.

1. There were numerous changes in production schedules for end-products made between the time the requirements were calculated and the quarter to which they were applicable. After 1942 these schedule revisions were usually decreases, rather than increases.

2. Spare parts requirements were usually calculated very generously to allow for high combat losses and damage. Experience has shown that these allowances were too high.

3. Scrap losses were actually lower than manufacturers calculated them to be, probably because of the increased skill and efficiency achieved in manufacturing operations.

4. Shortages of components caused industry to fall behind production schedules for end-products, and thus the aluminum which would have been required had the components been available was not ordered.

5. The human tendency to ask for just a little more than the minimum necessary, either for the sake of safety or for the sake of negotiation, was undoubtedly present, although it must be said in all fairness to the claimants that conscientious conservatism over-shadowed any apparent inflation of requirements for bargaining purposes.

Even if the claimants had been most successful in calculating their requirements, there undoubtedly would have been a small amount of attrition[6] caused by the administrative operations of CMP. The transfer of allotment papers through so many hands and the tendency to establish small reserves of allotments at each link in the chain would have, of itself, prevented the total use of allotments issued. Allotment paper followed the reverse course of requirements paper and passed from WPB to the claimant agency, to the prime contractor, to the subcontractor, and to the most remote sub-subcontractor, down the line. Attrition was recognized early as a concomitant of CMP and was compensated for by allotting a small excess over available supply.

The procedures and techniques used by the claimant agencies in their operations will make an important narrative that should be preserved for a future emergency. Perhaps the Army and the Navy will continue to calculate material requirements in the same way for peace-time programs, thus preserving a skeleton of their war-time materials planning organization. Although the ideal method of compiling requirements was described above, each claimant agency worked out modifications and devised substitutes for the bill of material procedure to meet its own circumstances.

EVALUATING CAPACITY AND FORECASTING PRODUCTION

One problem that defied complete solution and became a recurring source of difficulty in aluminum, and apparently in other materials planning was the determination of capacity. Perhaps no other aspect of WPB operations was featured by so much confused thinking and misunderstanding. Part of the explanation for this lies in the use of an unorganized system of nomenclature. The many terms used in this

[6] *Attrition* is a shop term denoting the difference between requirements, allotments, and orders placed.

field were never carefully defined and thus became crude tools of expression, vague in intent, implication, and interpretation. The use of different terms to represent the same thing and of the same term to represent different things is symptomatic of an effort to find a short-cut through a complicated subject but there is probably no short-cut to be found. We must recognize that in capacity studies, as in other statistical work, statistics are meaningful to the extent that (i) they are developed to serve a specific purpose, (ii) the assumptions on which they are based are clearly understood by both the statistician and his audience, and (iii) their probable error can be determined.

Capacity studies served two principal purposes in aluminum planning: first, to determine the need for building new plants to meet requirements on a long-range basis; second, to determine how much material could be produced in the facilities in operation to meet requirements on a short-range basis. In each case a plant-by-plant study was essential; indeed, for some commodities the study of the potentialities of each piece of operating equipment within a plant was called for. In studies initiated for long-range planning purposes the answer to this question was sought: How many pounds of commodity X can the aluminum industry produce monthly at peak with the pattern of business that can be forecast and with its present plants and the additional facilities already authorized or under construction, if there are no limitations on available manpower, production materials, maintenance, repair, and operating supplies, transportation facilities, and power? All the factors of production, in other words, are assumed to be at their optimum; the unknowns are the pattern of business inherent in the requirements and the potentialities of the physical production facilities in the light of the pattern. The reason for posing the question in this form is not obscure. There was no point in building new facilities if those in existence or already planned could, under the proper conditions, produce the amount of material required. If a shortage of labor, for example, were anticipated as a limiting factor in production, it had to be remedied directly. New facilities required so much construction time and critical materials that they could not be authorized if any other method of increasing production could be found.

The short-range study posed a different question: How many pounds of commodity X can the aluminum industry produce in a given quarter with the pattern of business and under the operating limitations with regard to manpower, raw materials, etc. that can be forecast for the quarter? (*Maximum practicable production* is the most descriptive term to represent this concept.) This question was easier to answer in some

respects, and more difficult in others, than the long-range question of supply—easier because it had to be answered only two or three months in advance of the quarter under consideration, but more difficult because of the many variables to be taken into account that were not susceptible of statistical treatment. The principal unknown was manpower. What urgency rating would the industry be granted for each of its plants? How effective would the rating be in increasing recruitment of labor by the United States Employment Service? Would interregional recruitment be permitted by the War Manpower Commission? What would be the impact of changes in Selective Service regulations? Would the Army furlough soldiers to work in the plants? Would the War Labor Board approve a pending wage increase? Would the National Housing Administration obtain needed trailer-housing in time? Would the Manhattan Project on one side of Knoxville continue to compete with the aluminum sheet mill on the other side? To what extent would seasonal factors affect labor supply? And how could the answers to all of these questions, if obtained, be translated into an estimate of manpower on the job and producing aluminum?

After the manpower uncertainties were resolved as far as they could be, the next point of investigation was the rate at which the industry could accelerate its production from the level at which it was expected to be operating immediately prior to the beginning of the quarter. Fluctuations in the demand for aluminum products were very marked during the last two years of the war, not only because of changes in procurement programs for end-products but also because of changes in the inventory policy of the aircraft industry. The year-end inventory clean-up drives of aircraft manufacturers caused aluminum orders and mill schedules to shrink in the last quarter of the year and rise again in the first quarter. This was particularly apparent at the end of 1944, before the reverses in the Ardennes, when the aircraft industry, with one eye on the headlines and the other on aircraft cutbacks, lived off their stocks and kept new orders at a minimum. In the face of such fluctuations in demand, the calculation of maximum practicable production for a quarter following one in which the mills were producing at relatively low levels, had to be tempered by the knowledge that building up maximum speed in an industrial plant was a gradual process. New workers, assuming that they could be hired, had to be trained and had to develop skills on the job, production lines had to be extended, new shifts established, working stocks built up, and so on.

The gathering of data for capacity studies was a joint task of the WPB production specialists and the statistical-planner. It was mani-

festly impossible for the latter (or his staff) to do the job without technical consultation. The statistician, however, laid down the ground rules so that uniformity would be achieved by the many people contributing to the task. The production specialists studied plant records, often visited the plants themselves, conferred with the producers, contacted other government agencies, and submitted a preliminary report. The statistician coordinated the data contained in all the reports filed, analyzed them in terms of past performance in comparable periods, brought to bear the manpower data available, and reported his recommendations to the executive in charge of the whole aluminum supply program. The commitments on supply were then made, usually after a final conference with the production specialists and the statistician.

The concept of maximum practicable production was not infrequently confused with *estimated production*. For example, one researcher proposed to measure the reliability of supply forecasts by comparing quarterly CMP supply commitments—the equivalent of maximum practicable production—with actual shipments. The results of such a comparison would be meaningless, for one variable that maximum practicable production figures did not take into account was the amount of orders that would be placed on the mills. Obviously the mills would not ship more than their customers ordered, with the possible exception of relatively small amounts preshipped against orders for delivery in future periods. The forecast of production (or shipments) as distinguished from maximum practicable production thus required a further leavening of the capacity concept by the introduction of another variable—order placement. Such forecasts were prepared by analyzing the requirements for a commodity and deflating or occasionally inflating them to realistic quantities to represent estimated order placement. When this figure was substantially lower than the maximum practicable production forecast, and it usually was, it was used to represent estimated production. Reports by industry of order acceptances for future months, while not a complete representation of the actual production load to be handled for more than one or two months ahead, were also of value in developing estimates of order placement.

The nature of the task of developing capacity and production estimates varied, of course, from product to product. Virgin aluminum is produced by an electrolytic process that requires continuous operation. Thus, the question of the number of shifts to be in operation during a period did not enter: a plant unit either operated full-time or it stopped operating. The pattern of business was not a factor; slight variations

may occur because of the changing needs to produce "rich alloys" in the reduction cells or pots, but the effect of this is negligible. A small slump usually occurs in the summer months, but otherwise, excluding curtailments caused by policy decisions or severe labor or power shortages, production is very stable. Secondary metal is in a class by itself. While the calculation of smelter capacity and maximum practicable production present no unusual problems, the forecast of smelter production and of net secondary recovery is a complicated affair in which the element of the informed guess must enter strongly. Smelter production is influenced by the relative prices of primary ingot, secondary ingot, and aluminum scrap, with the latter being particularly subject to rapid market fluctuations. The policies of integrated producers which use large quantities of scrap also influence secondary recovery substantially. The manufacture of semi-fabricated products is subject to all possible influences, including the pattern of business, the number of shifts that can be operated, the heat of operation in the summer (notably in foundry work) as well as labor supply, the supply of fuel for heating melting furnaces, and the like.

THE CONCEPT OF "NET SHIPMENTS"

Aluminum may pass through several stages of fabrication and through the hands of a distributor before it is shipped out of the industry to the industrial consumer.[7] In this respect aluminum is no different from other metals which serve structural purposes. While ingot shipments might be used as the basis for a statistical system in aluminum, such a measure, only indirectly responsive to changes in industrial demand, would be of limited value. Measuring aluminum activity at the semi-fabricated product level is much to be preferred, although it involves several problems in surveying the industry and interpreting the tabulated data.

It is plain that if we were to add together shipments made by all plants in the industry our total would be inflated to the extent that the same metal is counted as a shipment more than once. Rivets shipped by a distributor to a cooking utensil manufacturer, to take an extreme case, would be counted as ingot, rod, wire, and rivets. And if distributors were included in the system, the rivets would be counted twice as rivets: once when shipped by the rivet manufacturer to the distributor and once when shipped by the distributor to the final consumer. It was apparent at the start of planning the aluminum program that, while it

[7] A relatively small amount of aluminum is used outside of the aluminum industry as ingot; for example, in deoxidizing steel and alloying magnesium.

was essential to know how much of each aluminum product was produced and shipped, regardless of use, it was no less essential to know how much aluminum was being obtained by military contractors and other consumers. Neither gross nor net figures could be dispensed with.

The problem of calculating net shipments was solved by assigning each producer and distributor in the industry with an identification number which he was required to endorse on the aluminum orders he placed with other producers or distributors. This requirement was applicable also to the separate departments of the integrated producers, each department being treated as a separate unit. With this sort of identification on orders received, it was possible for each aluminum supplier to report his shipments divided among shipments to other producers (called shipments for further fabrication), shipments to distributors, and shipments to ultimate consumers. The assignment of identification numbers was arranged schematically so that each number not only indicated that the enterprise using it was in the aluminum industry but also distinguished a producer from a distributor and, in the case of a producer, defined the type of aluminum product he would fabricate from the aluminum product shipped to him.[8] A technique somewhat similar to this was used by WPB during the war for the steel industry. Steel used a uniform symbol rather than an individual number, however, and did not include in gross shipments, transfers of material among the separate departments of an integrated company.

INDUSTRY STOCKS AND THE "PIPELINE"

Closely allied with concept of net shipments is the concept of "pipeline." The pipeline of the aluminum industry may be defined as the amount of ingot, in-process metal, and semi-fabricated products held by, and material in transit to, plants in the industry that is required to enable the industry to ship its products at a given level. A brief equivalent might be "minimum working stocks."

It is a happy coincidence that the processing losses encountered in converting aluminum pig to semi-fabricated products closely approximate the weight of ingredients (copper, magnesium, chromium, etc.) added in the production of aluminum-base alloys. Thus the net change in the industry's total working stocks during a period can be estimated within narrow limits by subtracting net shipments of aluminum products from new supply (virgin metal production, net secondary recovery,

[8] This method of identification proved particularly valuable as it provided WPB with the data for determining the quantity of one aluminum product required to produce a unit of another; for instance, the quantity of tube blooms required per pound of tubing. Reliable ratios of this kind were essential tools in all WPB metals research.

plus imports) for the period and making an appropriate adjustment for any change in government-owned stocks. This principle becomes more useful when it is combined with another, namely, that the industry requires two and eight-tenths to three pounds of aluminum in its pipeline to make *monthly* net shipments of one pound of aluminum products.

Once a base inventory figure was established it was possible, with these principles and without a series of inventory reports, to determine the stock position of the industry at the end of each month and particularly to observe the trend of industry stocks over a series of months. It was also possible to forecast with great facility the amount of metal that the industry itself would require for the pipeline (that is, over and above the amount to be shipped to consumers) to meet increased total requirements, and, conversely, the amount of metal that could be made available from the pipeline if requirements fell.

The following formulae will illustrate:

$$P_c = V + S_n + I - N + (G_b - G_e) \tag{1}$$

$$P_c = 2.8\,\frac{R_q - R_{q-1}}{3} \tag{2}$$

where

P_c = pipeline change
V = virgin metal production
S_n = net secondary recovery
I = imports
N = net shipments of aluminum products
G_b = government-owned stocks, beginning of period
G_e = government-owned stocks, end of period
R_q = requirements for a designated calendar quarter
R_{q-1} = requirements for the preceding calendar quarter

Data to solve these equations were available from reports filed by industry or government agencies. The change in the pipeline for a past period could be calculated through equation (1); the change in the pipeline for a future quarter could be forecast through (2).

From this discussion it might appear that it was assumed that the aluminum holdings of the industry were always at the minimum required levels and that the pipeline changes indicated by the preceding calculations would occur immediately. It was recognized, however, that changes in demand sometimes were too rapid and too extensive to permit the industry to adjust its stock position in a month or two. In addition, domestic virgin aluminum production could not be curtailed as

rapidly as requirements fell because of the possibility that subsequent increases would call for reopening facilities, a difficult accomplishment once the labor supply was absorbed by other industries. Since production curtailments generally lagged behind decreases in requirements and since the government did not purchase domestically-produced virgin aluminum for stockpile purposes, the producers of virgin metal carried for a substantial period of time much more metal than they required for their operations. The excess stocks held by the industry (E) were estimated in consultation with the producers, and while excess aluminum industry stocks existed, (1) became:

$$P_e = V + S_n + I - N + (G_b - G_e) + (E_b - E_e). \qquad (3)$$

It might also appear that the excess stocks of the industry could be determined from monthly net shipments data, as follows:

$$E = P - 2.8N \qquad (4)$$

where P represents the total pipeline calculated by adding the monthly pipeline increments to the base inventory. Reliance on this formula would have been tantamount to assuming that the processing losses *exactly* equalled alloying ingredients, for even a small monthly error over a period of several years would accumulate to large proportions. This assumption could not be made with confidence,[9] and accordingly other techniques were used to estimate the excess metal that existed at any time.

The usefulness of the pipeline formulae made it unnecessary to obtain monthly inventory reports from the whole aluminum industry. Regular inventory reports were obtained from the primary metal producers and the secondary smelters, but the rest of the industry was free of inventory reporting except when some problem requiring a knowledge of the location of specific products or alloys made special reports necessary.

SECONDARY METAL AND SCRAP

Secondary metal supply is another subject which had to be cut to the measure of war-time statistical requirements. Fortunately the pre-war surveys and analyses of the Bureau of Mines provided an excellent starting point for the WPB program in this field. The Bureau received

[9] In fact, a review of the monthly ratios of P/N showed a steady rise that could not be satisfactorily accounted for. This indicated a small monthly discrepancy cumulating over almost four years which inflated substantially the total amount calculated to be in the pipeline. This tendency to overstate the pipeline was not great enough, however, to impair the usefulness of the formulae for forecasting pipeline changes six or nine months ahead.

reports on all transfers of scrap *between companies*. The resulting recovery figures therefore encompassed new scrap generated in the aluminum industry, new scrap generated by finished product manufacturers, and old scrap;[10] they did not encompass new scrap that was used within the company in which it was generated. Thus, at least implicitly, the Bureau defined secondary recovery as metal recovered from scrap used in a company other than the one which generated it.

There is no argument that such a definition is satisfactory for ordinary purposes inasmuch as the commercial aspect of the secondary metals industry is as important as any other in peace-time. Nevertheless, an analysis of this approach in terms of the requisites of the wartime statistical program for aluminum showed that new principles had to be established, as follows:

> First: Secondary recovery data must *exclude* metal recovered from new scrap generated in the course of producing any semi-fabricated aluminum product within the defined scope of the aluminum industry, even if such scrap is sold by one company to another.
>
> Second: Secondary recovery data must *include* metal recovered by the aluminum industry from new scrap generated in the course of consuming (for example, machining) semi-fabricated aluminum products, even if the scrap is generated by a company which remelts such scrap itself and does not sell it to others.
>
> Third: Secondary recovery data must include all metal recovered by the aluminum industry from old scrap.

The first principle calls for the exclusion of certain types of scrap included by the Bureau of Mines, namely, scrap generated in the aluminum industry and sold by one company to another. The metal embodied in scrap of this type was never incorporated in a semi-fabricated product shipped by the industry. It was counted as an addition to supply when originally produced as virgin metal and cannot be counted again as secondary metal until it passes out of the industry as an item included in net shipments, and then returns. It must be remembered that the aluminum industry includes semi-fabricated product manufacturers as well as ingot makers and that the WPB balance sheet of the industry was designed to measure unfabricated supply on one side and semi-fabricated shipments on the other. All the metal between these two points of measurement constitutes the pipeline, and until metal leaves the pipeline it cannot be counted as an addition to supply in secondary form.

The second principle calls for the inclusion of certain types of scrap

[10] *Old scrap* is scrap obtained from obsolete, broken, or worn-out finished products, such as wrecked automobiles. *New scrap* is scrap generated in the course of manufacturing operations.

excluded by the Bureau of Mines, namely, scrap generated in the course of the ultimate consumption of semi-fabricated products but remelted by the company which generates it. Scrap of this type, transferred within a company which is integrated to the extent that it both produces and consumes aluminum products, is no different from scrap generated by a non-integrated consumer. To illustrate: Company A manufactures cooking utensils and the aluminum sheet from which they are made. Scrap generated in the cooking utensil department is sent to the sheet department for remelting and rerolling into sheet. The sheet is then sent back to the cooking utensil department for use. Company B manufactures cooking utensils exclusively, using sheet purchased from Company C. Scrap generated by B is shipped under a toll agreement to C for remelting and rerolling into sheet, which is then returned to B. Company D also manufactures cooking utensils exclusively, using sheet purchased from C. But scrap generated by D is sold to Company E, a smelter. Company E remelts the scrap and sells ingot to Company F, which rolls it into sheet for sale to Company G, an aircraft manufacturer. Treatment of the scrap in all these three cases must be the same in materials calculations. Although the scrap moves through three different channels on three different commercial bases—intra-company transfer, toll, and sale—the material movement and the contribution to supply in the form of secondary recovery are identical for the aluminum industry taken as a unit.

There is no difference between the WPB and Bureau of Mines secondary recovery calculations with respect to the third principle dealing with the treatment of old scrap. There is one other point of difference, however. The Bureau includes aluminum scrap used outside of the aluminum industry, for example, in the production of chemicals. Since WPB was comparing the supply available from, and the requirements placed on, the aluminum industry and since demands of other industries for aluminum scrap were not included in aluminum requirements calculations, the use of aluminum scrap outside of the aluminum industry was not included by WPB in the calculation of net secondary recovery.

The theoretical development of these principles was a much more simple task than their application to the collection of secondary metal statistics and the calculation of net secondary recovery. To achieve perfection it would have been virtually necessary to put a tag on each piece of scrap so that after it moved from generator, to dealer, to user, its identity could be determined and reported. Fortunately an analysis of aluminum production and scrap trade practices revealed that a re-

porting system, not markedly different from that used before the war by the Bureau of Mines, would provide the basic data which, when adjusted to eliminate double-counting of new scrap generated and sold within the aluminum industry, would give a secondary recovery figure representative of the net addition to the aluminum supply obtained from scrap.[11] The WPB calculation proceeded along the following lines. Recovery of scrap by smelters was calculated by adjusting their total ingot production figures to eliminate that portion of the ingot obtained from materials blended with the scrap (virgin metal, secondary pig, and alloying ingredients). The smelters figures were also adjusted to remove the amount recovered from scrap processed on toll. This type of scrap is usually generated by aluminum foundries; hence it is considered the equivalent of "run-around" scrap circulating within the industry even though sold by its generator. Scrap purchases of integrated producers of virgin metal, including scrap received from their end-product departments, less an estimated melt loss, were also included. Receipts, less estimated melt loss, rather than actual consumption and recovery, were used in this case primarily because of convenience in reporting. Scrap receipts of foundries and other producers of semi-fabricated products were included in similar fashion. From the total of these three items was subtracted the estimated quantity of scrap generated by producers of semi-fabricated products who did not have melting facilities, i.e. the producers who ship all of their scrap to others within the industry. This estimate was calculated by multiplying the reported shipments of semi-fabricated products made by these producers by factors obtained from industry experts representing the amount of scrap generated per pound of product manufactured. A formula summarizing the whole calculation would read as follows:

$$S_n = S_s + S_i + S_f - (.26F + .08R + .70T) \qquad (5)$$

where

S_s = secondary recovery by smelters, excluding recovery from their own "run-around" scrap and scrap processed on toll.

S_i = secondary recovery by integrated metal producers, including recovery from scrap generated in their own end-product departments, but excluding recovery from their own "run-around" scrap.

S_f = secondary recovery by producers of semi-fabricated products, including recovery from scrap generated in their own end-product departments, but excluding recovery from their own "run-around" scrap.

[11] Scrap data were obtained from aluminum scrap dealers through [illegible] smelters in 1945 by the Bureau of Mines as agent for WPB; smelters data [illegible] scrap data were obtained by WPB itself.

F = shipments of forgings by producers without melting facilities.
R = shipments of rivets by producers without melting facilities.
T = shipments of tubing by producers without melting facilities.

FOREIGN TRADE STATISTICS

Both for reasons of speed and convenience and because of the nature of the Controlled Materials Plan the data developed in WPB on aluminum imports and exports are noticeably different from those compiled by the Department of Commerce. The discrepancies over the period 1941–1945 taken in the aggregate should be negligible, however, for the differences are attributable principally to timing rather than method of measurement.

Beginning in 1941, the United States, through the Office of Metals Reserve, made a series of contracts calling for over a billion pounds of aluminum to be delivered by the Aluminum Company of Canada. Transactions under these contracts were entered into WPB import statistics as soon as the aluminum was delivered to warehouses in Canada for U. S. account or, in the case of metal sent directly to the U. S. by the producer, at the time it was shipped by rail. This method of calculation was used because U. S. holdings in Canada were included in U. S. stockpile figures, and the aluminum metal accounting system would have been out of balance otherwise. It should be noted that not all of the metal contractually delivered crossed (or will cross) the border. Some of it was resold to Canada, some was transferred to the account of the U. K. in repayment for loans of aluminum made to the U. S. early in the war, and some was sold to the U. K.

Imports of relatively small quantities of aluminum were also made under other special arrangements. The U. S. received aluminum semi-fabricated products of Canadian manufacture for which we provided the aluminum raw materials. On the other hand we received ingot from the U. K. in reimbursement for various aluminum semi-fabricated products which we provided through the Foreign Economic Administration and in reimbursement for aluminum powder provided through the Navy Department.

Reports of shipments filed by aluminum producers and distributors were used in compiling export data. As far as CMP was concerned the job of supplying a requirement was completed when the aluminum industry shipped the materials ordered out of its plants. Accordingly no regular effort was required or made to follow up mill and warehouse shipments to the point of actual export. WPB export data therefore reflect the time of mill shipment rather than the time of actual export.

These data include Lend-Lease shipments as well as shipments on direct sale; they do not include material purchased by the Army and Navy and destined ultimately for shipment for their own use abroad or for the use of a foreign country under the International Aid program. Department of Commerce figures are similarly compiled, except with respect to timing and except for International Aid shipments, which are included in the official foreign trade statistics.

* * *

The experience of the War Production Board has several significant implications for the future. It showed, if further proof were needed, that "descriptive" and "mathematical" statistics are indispensable tools not only of governmental but also of industrial administration and planning. It showed that the nation's springboard for total production is a detailed knowledge of the condition of its industrial machine and that the pre-war surveys of the Federal government were not good enough for this purpose. Finally, it showed that a successful governmental statistical program must be a coordinated program, molded by a constructive and central force. The consideration already given to the integration of Federal statistics[12] is a hopeful indication that the handwriting on the wall is being read by statisticians.

[12] See "Problems of Integrating Federal Statistics: A Round Table," *Jour. Amer. Stat. Assoc.*, Vol. 40, No. 230, June 1945, pp. 237–244.

HISTORICAL NOTE ON THE PURCHASING POWER CONCEPT, AND INDEX NUMBERS

WIRTH F. FERGER
Bureau of Internal Revenue
U. S. Treasury Dept.

IN HIS historical appendix to *The Making of Index Numbers*, Irving Fisher credits to Dutot, in 1738, the construction of the first index number of prices, a simple aggregative. A few years ago I picked up in a second-hand book store in Columbia, S. C., a book published in London in 1707 in which the author computes a kind of mental weighted average of relatives index of retail price changes between 1440–1460 and 1686–1706. Whether the arithmetical application of the formula is sufficiently precise to give the author credit for the first index number is open to question, as will be seen, but there can be no doubt that he had a clear concept, not only of the purchasing power of money and its significance, but also of the principles of constructing both weighted aggregative and weighted average-of-relatives index numbers which the succeeding 240 years have only refined, not changed basically.

The title page does not disclose the name of the author, but bibliographers credit the book to William Fleetwood, the Bishop of Ely. The title page reads as follows: "*Chronicon Preciosum:* or, an ACCOUNT of ENGLISH MONEY, the PRICE of CORN, and *Other Commodities*, For the last 600 Years.—in a LETTER to a *Student* in the University of *Oxford*.—LONDON, Printed for Charles Harver, at the *Flower-de-luce*, over-against St. *Dunstan's* Church, in *Fleetstreet*. MDCCVII." Two reprints appeared in 1737 and 1745. The book has six chapters, one given to a statement of the problem, two to accounts of gold and silver coins, two to scattered quotations of commodity prices and wages "for 600 years last past," and a final chapter giving the conclusion. Present interest centers on the first and last chapters.

It is interesting to observe that a practical problem of morals gave rise to the speculations of the author—or did it really? One is led to wonder if the Bishop, having spent a good deal of time looking into the history of coins, prices and wages, did not invent a practical question as an expositional device—not a blame-worthy motive. Be that as it may, he also found a second and a third application of his theory in his Conclusion. Let the Bishop state the proposition, as submitted by the "student":

"CHAP. I.

"*The* CASE

"The Statutes of a certain College (to the Observation of which, every one is sworn, when admitted Fellow) vacating a Fellowship, if the Fellow has

an Estate in *Land of Inheritance*, or a *perpetual Pension*, of *Five Pounds per An.* I desire you would be pleased, to give me your Answer to these following Questions; when I have first told you, that the College was founded between the Years 1440, and 1460.

"I. Whether He who is possessed of an Estate, of that, or greater Value, may make it over *in trust* to his Friend, and then safely swear to the Observation of such Statute, amongst the rest?

"II. Whether He who has not an Estate of that Value, when admitted Fellow, may keep his Fellowship tho' he *afterwards* come to an Estate of that, or much greater Value?

"III. Whether He who is actually possessed of an Estate, of *Six Pounds per An.* as Money, and Things go *now*, may safely take that Oath, upon presumption, that VI *l.* now, is not worth what V *l.* was, when that Statute was made."

Bishop Fleetwood found no difficulty in coming to a negative answer on the first two questions. Only the third question disturbed the Bishop and interests us:

"Your *Third* and last Question will cost me more Pains, and you more Patience, before we come to the Conclusion; if we can come to any satisfactory one at last. The Question would certainly need no Answer, if it were asked in gross, *Whether he who has VI Pounds per An. may safely swear he has not V?* When a Pound is, in both Places (and has been so for more than 600 Years) XX *s.* but as you have qualified it, by distinguishing the Times, it will require both a good Casuist, and a pretty good Historian, to answer it absolutely, and to your Purpose; which is (as I take it) to know, of what Value you may now hold an Estate, with safety to your Conscience, which is charged with the Observation of that Statute, which vacates the Fellowship of One who has an Estate of Inheritance, or a perpetual Pension of V *l.* or VI *l.* *per An.* A better Casuist, I own, you might easily have found. But, it may be, you could not so easily have found One, who hath in his Reading, made more Observations on the Price of Corn, and other Commodities, at different Times, than I have done, as you will perceive by reading this long Letter. So that, for ought I know, my Diligence may make you Amends for what want of Judgment may be found in me. And I do not say this in Vanity and Ostentation of my Labour, but because without a good share of Knowledge in these Matters, the best Casuist in the World can never answer your Question satisfactorily. . . . "

The Bishop need not have so disparaged his own "casuistry," for after a few remarks about the "good use, in the Consideration of great Affairs" that may come from "Collections of so mean Things, as the Price of Wheat, and Oats, of Poultry, and such like Provisions," he argues the case in a manner that would do credit to an Economic Statistician a century later.

"I do affirm then, with the best Judgment I have, that I am sincerely persuaded, that, altho' you are actually possessed of an Estate of VI Pounds

per An. as Money and Things go now, you may safely take that Oath, upon Presumption that VI Pound *now*, is not worth what V *l.* was then, when that Statute was first made. Because whoever swears, swears to *Things* that are signified by Words, and not to *mere words*. When a Word signifies the same thing *now* in effect, which it signified 200 Years ago, then he who swears to *Words*, swears to the Things they signifie; but when different Things are signified by the same Word, then he who knows that difference of Things, cannot help giving such Word, its proper and intended signification. A *Pound* (for instance) will buy either more, or less Corn (take it which way you will) *now*, than it would in *H.* VI [Henry the Sixth's] time. A *Pound* is therefore of more or less Value *now*, than it was *then*; and the *Value* of a *Pound* is truly a *Pound*, and not its mere *Name*. It is not therefore the same *Thing* now, that it was in *H.* VI. time."

Have the succeeding 238 years brought a better statement of the concept of the purchasing power of money? The author even had a clear notion of a regimen of commodities to be used in a weighted aggregative index number:

"And that I may very honestly have Regard to the Value of V *l.* 260 Years ago, will, I believe, appear evident from what I am going to say; That the Founder intended the same Ease, and Favour to Those who should live in his College 200 Years after his Decease, as to Those who lived in his own Time. Now, they who lived in his Time, might, with V Pounds, purchase so much Bread, so much Drink, Meat, Cloth, Firing, Books, and other Necessaries, or Conveniencies: I know not exactly how much, nor is it material: I only say, the Founder intended I might keep such an Estate, as would suffice to procure the same Bread, Drink, Meat, Cloth, Books, &c. as the other might have procured for V Pounds, 260 Years ago. But this I cannot possibly do with VI Pounds, as Things go now, nor it may be, with four times as much. . . . This is a clear Proof, that Regard both may and must be had, to the different Value of Money, at different Times; and that the Founder's visible Intention is better answered by such Regard, than it would be by a strict and obstinate Adherence to the bare Letter of the Statute."

The author goes on to say that it is not his purpose to determine exactly to what extent prices have risen, or the value of the pound has changed: "I leave that to Others, to gather from the particular Accounts of Things that I shall give them, from our Historians." But he was sure that although a pound in the time of Henry the Sixth might be worth 4 or 5, or more, in his own time, it was certainly not worth 20 or 30. This leads to the chapters on the description of gold and silver coins, to serve as the basis of the prices quoted in the succeeding chapters. Chapter IV. "Of the Price of Corn, and other Commodities, for 600 Years Last Past," is introduced as follows:

"Having dispatch'd, with what *exactness* I could, the Chapters of *Moneys*, I am now come to speak to the Price of *Corn* and other *Commodities*; which

Is (whether you know it, or not) the readiest way to the solution of your *Third*, and most material Question. For, your Business is to know (as near as you can) what Estate or Summ of Money will *now-a-days* be equal, or equivalent to *five Pounds* (let that be the supposed Summ in this Discourse) in the Reign of King *H.* VI. and to this End, your Care will be, to find out how much *Meat, Drink*, or *Cloth*, might be purchased in *H.* VI. Reign, with V *l.* and then to find out, how much of the Money now current, will be required to purchase the same quantity of *Meat, Drink*, and *Cloth*. For, since Money is of no other use, than as it is the Thing with which we purchase the Necessaries and Conveniencies of Life, 'tis evident, that if V *l.* in *H.* VI. Days, would purchase 5 *Quarter of Wheat*, 4 *Hogsheads of Beer*, and 6 *Yards of Cloth*, he who then had V *l.* in his Pocket, was full as rich a Man as he who has now XX *l.* if with that XX *l.* he can purchase no more *Wheat, Beer*, or *Cloth*, than the other. I do not mean hereby to pre-judge this to be the Proportion; but use this Instance to let you see, that this is the properest way of coming to know, what Estate is *now* most answerable, to an Estate of V *l.* *per An.* 250 Years ago. . . ."

Curiously enough, this concept of a weighted aggregative index number, as we call it, became transmuted into an average-of-relatives in Chapter VI, the Conclusion, for the price increase of each of several sample commodities is estimated and a mental average taken, to express the author's final judgment. After explaining the difference in the number of shillings in a pound between the two dates under discussion, the Bishop suggests that instead of taking one particular year at the present and 250 years earlier, "you must take the Price of every particular Commodity, for as many Years as you can (20, if you have them) and put them all together; and then find out the common Price; and afterwards take the same Course with the Price of Things, for these last 20 Years; and see what Proportion they will bear to one another; for that Proportion is to be your Rule and Guide." The succeeding paragraph illustrates the method by assuming certain prices for wheat, oats, beans, ale, and cloth ("Good *Cloth*, such as was to serve the best *Doctor* in your *University*, for his Gown"); but from the language it appears that the Bishop believed that he was using realistic prices. Wheat, oats and beans were found to have increased in price by about six-fold over the 200 years. "But you must not expect that every Thing will answer thus exactly." Ale and cloth had risen about five-fold in price, and he had good reason to believe that 200 years earlier beef, mutton, bacon "and other common Provisions of Life, were six times as cheap." The final statement is as follows:

"And therefore I can see no Cause, why 28, or 30 *l.* *per An.* should now be accounted, a greater Estate, than V *l.* was heretofore, betwixt 1440, and 1460."

Two other applications of the price index concept are mentioned by the author, the first of which can certainly be regarded as important,—in fact a matter of life and death! He quotes Sir *H. Spelman*, "a very competent Judge and Aestimator of these Matters," to the effect that the laws providing capital punishment for conviction for theft of things to the value of twelve pence were being invoked to condemn to death many people whose crime would not have brought the death penalty when the laws were enacted many years earlier. Finally, and to end up on a less gruesome note, the cause of democracy was served by inflation, when the qualification for "*Electors* of Parliament-Men" had been set years before at the possession of lands or tenements to the yearly value of 40 shillings.

> "Whether the Legislative Power, in 1430, did well, and wisely, in reducing the Number of Electors to such as were worth 40 *s. per Annum*, (which cut off many hundred thousand voices, and consequently many occasions of Tumults and Disorders) is not to be doubted over-much; nor yet is to be over-confidently affirmed, because if it had been so wise and useful an Ordinance, it would have still been kept up, in its due Proportion, according to the difference of Times; altho' the Changes of such Moment are not to be frequently and lightly made. But in these Affairs, it is not fit for private People to meddle."

A final bit of advice was given by the Bishop to his correspondent:

> ". . . And that is, That if ever you design to take Orders, and obtain any *Rectory*, *Vicarage*, or higher *Dignity* in the Church, you be, above all Things, careful, how you make any Composition or Agreement, for any long Space of Years, to receive a certain Price of Money, for the Corn that is due to you, altho' for the present it may seem a tempting Bargain, and a profitable Exchange, and rid you of some Trouble, You know not what Time may bring forth, nor what great Alterations may happen, nor what great Mischiefs you, unwittingly, may do your Successors. . . . "

A METHOD OF MAKING ACTUARIAL ESTIMATES FOR A COMPULSORY HEALTH INSURANCE SYSTEM

PAUL A. DODD, Ph.D.,
Professor of Economics
University of California

THE actuarial basis of a compulsory health insurance system, in the final analysis, resolves itself into a problem of simple arithmetic. Obviously enough money must be raised under the system to pay for the benefits provided thereby, and benefits provided under the plan cannot be more extensive and complete than what can be supported by the health insurance fund. In other words, under a compulsory health insurance plan, as with all other things, it is impossible to get something for nothing, and all that is offered under the insurance system must be paid for out of funds collected from the employer, employee, and/or the government.

There are at least two approaches to the problem of determining the costs of such a sytem. One may fix arbitrarily the maximum amount of contributions that it appears wise or prudent to levy for the plan, thereby determining the amount of money with which services can be bought. Or one may determine what appears to be the minimum number of basic services which should be offered in bringing basic medical care to the desired number of persons, and then, upon a basis of these needed services, determine how much money will be needed to pay a fair price for those services. In either case it will be necessary to estimate as accurately as possible (a) the total or per capita cost of basic services to be afforded, and (b) the total or per capita contribution which will be collected to pay for the services.

Since compulsory health insurance is a new field to the people of this country, both the amount of the levy which can and should be collected for the support of the scheme and the number and nature of what are to be called "basic medical services" afforded to the largest possible number of the population are matters that, for the want of a better method, must be more or less arbitrarily determined. But a start is needed, and enough services under the health insurance system must be offered to meet at least the basic needs of the largest number of persons possible at a cost that will not be over-burdensome for the individual, the employer, or the state or federal government. Otherwise, the plan cannot survive, and cannot be instrumental in meeting the basic difficulties encountered under the system of the private practice of medicine.

Several compulsory health insurance plans were presented at the last session of the California State Legislature. Even though no comprehensive compulsory program has yet been envoked in any state of this country, in view of the widespread application of the health insurance principle in most other advanced nations of the world, and the imminence of its application either upon a state-wide or nation-wide basis in the United States, it seems clear that a sober and sound approach to the actuarial basis of a compulsory plan is already long past due. Inasmuch as an American plan for compulsory health insurance must not only profit from the mistakes that have been made in this form of social insurance abroad, but also must be based upon American experience and American standards, an approach to this problem has little to build upon as a foundation.

In order that a start may be made in the development of a sound actuarial basis for such a program, this discussion will suggest a method of approaching an answer to the question as to how much revenue could be expected to arise under a compulsory health insurance program such as that proposed for California by Governor Warren early in 1945. Several basic computations are involved in the development of a method:

1. Estimates of taxable earnings and tax revenues, by years, during the ten-year period 1938–1947.
2. Estimates of the number of employees and dependents in California who would be brought within the scope of the system outlined in the Governor's proposal (Assembly Bill No. 800), or its counterpart, and introduced in the State Legislature early in 1945.
3. Estimates of the average per capita amounts made available through the health insurance fund and covering, in the main, approximately the same working groups as are now covered under the unemployment compensation law, but including dependent family members as well.

A fourth basic computation, that of estimating the amount of medical benefits that can be offered under the proposed program, cannot be presented within the confines of this paper.

I

Estimates of Taxable Earnings and Tax Revenues in California 1938–1947

Any sound approach to the actuarial basis of a compulsory health insurance system must start with the Federal Census of Population and an analysis of the numbers of persons gainfully employed. The number, sex and age distribution of employees to be covered by the compulsory provisions of a health insurance plan for California have been made as

of January 1, 1935[1] and for 1937,[2] and the method developed need not be re-presented here. It will, of course, be necessary to project population trends for the state from the last Federal Census period (1940) for several years. As indicated below in Table 4, this study projects these trends through the year 1947, at which time it is estimated the state will have a total population of 8,440,000 resident civilians.

The age and sex distribution of the population, and the major occupational groupings of the population are easily obtainable from the

TABLE 1

INCOME PAYMENTS IN CALIFORNIA COMPARED WITH VARIOUS SHARES OF TOTAL INCOME PAYMENTS AND AVERAGE PER CAPITA CIVILIAN INCOME PAYMENTS AFTER FEDERAL TAXES, 1935–1947 (ESTIMATED FROM 1944)

	Total Income Pmts. all persons (M-$)	Income Pmts. to civilians (M-$)	Total Salaries and Wages (M-$)	Total Net Income to Proprietors (M-$)	Dividends, Interest and Net Rent (M-$)	Other Income Pmts. (M-$)	Total Civilian Income Pmts. after Fed. Income Taxes†	Per Cap. Pmts. after Taxes
1935*	$ 3,952	$ 3,052					$ 3,[illegible]	$ 650
1936*	4,780	4,780					4,[illegible]	773
1937*	5,105	5,105					5,[illegible]	729
1938*	4,808	4,808					4,[illegible]	[illegible]
1939	5,047	5,047	3,050	750	[illegible]	[illegible]	4,[illegible]	[illegible]
1940	5,605	5,605	3,389	875	954	[illegible]	[illegible]	[illegible]
1941	7,044	7,000	4,310	1,150	1,190	350	[illegible]	[illegible]
1942*	9,260	8,960					[illegible]	1,[illegible]
1943	12,097	11,318	7,810	1,798	1,325	[illegible]	10,[illegible]	1,[illegible]
1944	—	12,300	8,350	2,000	1,450	500	11,[illegible]	1,[illegible]
1945	—	10,500	6,400	3,500		[illegible]	[illegible]	1,[illegible]
1946	—	9,150	5,450	2,000		[illegible]	[illegible]	[illegible]
1947	—	9,400	5,650	2,950		[illegible]	[illegible]	[illegible]

* Detailed estimates not available.

† Allowing for a 10 per cent Federal Income tax deduction for the years 1945–1947.

Source: Totals for 1935–1943 inclusive from United States Department of Commerce Reports on Income Payments, by States. Other amounts (years 1944–1947) are estimates based upon projected employment and postwar income studies made by the State Reconstruction and Reemployment Commission, and the Brookings Institution.

Federal Census reports, and require little if any adjustments to provide estimates for the several years in the immediate future for which data must be made available.

Thus the launching point in our study is reached in connection with estimating the amounts of total and per capita income payments, for such data must serve as a basis for estimating the total revenues to be

[1] Dodd, Paul A. and E. F. Penrose, *Economic Aspects of Medical Services, with Special Reference to California*, Graphic Arts Press. (Washington, D.C.) 1939, pp. 415–458.

[2] *Estimating the Costs of Medical Benefits Under a Proposed Compulsory Health Insurance System for California*, by Paul A. Dodd, University of California, April, 1940 (Mimeographed), p. 4.

raised by means of a specified payroll tax schedule as stipulated in the health insurance proposal. These data are presented in Table 1, where phenomenal increases in total income payments received by persons in California between 1935 and 1947 are indicated. In 1935, for instance, persons in California received an estimated total income of $3,952,000,000, or an average per capita income, after federal taxes, of $650. By 1943 this estimated total had increased to over $12,000,-000,000, or to an average after federal taxes of $1,335 per capita.

TABLE 2

NUMBER OF EMPLOYEES WHO RECEIVED WAGES IN COVERED EMPLOYMENT (UNEMPLOYMENT INSURANCE ACT) DURING 1940 AND 1943 IN CALIFORNIA

Wage Group	Number and Percentage of Employees			
	1940		1943*	
	Number (in 000s)	Per Cent	Number (in 000s)	Per Cent
$ 1 to 199	570	26.4	928	24.4
200 to 399	254	11.6	400	10.5
400 to 599	180	8.6	280	7.4
600 to 799	140	6.8	213	5.5
800 to 999	152	6.9	176	4.6
1000 to 1499	341	15.5	409	10.8
1500 to 1999	262	12.0	344	9.0
2000 to 2499	144	6.6	291	7.6
2500 to 2999	57	2.7	276	7.3
3000 and over	64	2.9	483	12.7
3000 to 3999†	10	.85	145	3.8
4000 to 4999†	10	.85	145	3.8
5000 and over†	26	1.2	193	5.1
Total	2,191	100.	3,800	100.

* Wages for 1943 represent wages until September 10, 1943.
† Estimated distribution made by the C.E.S.C.
Source: "*A Report of Activities for the Month of January, 1945*," California Employment Stabilization Committee.

The most reliable estimates obtainable indicate that in 1940 total "salaries and wages" in California approximated some $3,389,000,000; by 1943 they had exceeded $7,800,000,000; and in 1944 they had reached a peak of some $8,350,000,000. This latter year the estimated average per capita income payments (after federal taxes) also reached a peak of $1,405. These data also suggest that while the total "wages and salaries" in California had passed their peak with the passing of 1944, still, by 1947 they will probably approach an estimated total of some $5,650,000,000, an amount 65 per cent above the level attained in 1940.

In Table 2 are presented basic data relative to the distribution of earnings among employed workers who received wages in covered employment in California during 1940 and 1943. Since most compulsory health insurance proposals are advanced under the assumption that some minimum earnings level must be applied to establish eligibil-

TABLE 3

TOTAL WAGES, TAXABLE WAGES, AND ESTIMATED TOTAL TAXABLE WAGES UNDER THE WARREN HEALTH INSURANCE PROPOSAL

Estimates Based Upon Employment Covered by the California Unemployment Insurance Act, 1938–1944 (Including Estimates for 1945–1947)

Year	Total Wages† in employment covered by the C.U.I. Act	Taxable Wages subject to the Cal. Unemp. Insurance Act ($3000 limit)	Actual U.C. Tax (rate 2.7% + 1.0%) into the Cal. Unemp. Ins. Trust Fund	Est. Rate of U.I. Tax based on 3.7% ($3000 limit)‡	Est. Total Taxable Wages and Salaries of Health Plan (Warren)§	Est. Rate of U.I. Tax based on 3% tax
1938	$2,004,178,555	$2,004,178,555	$ 82,300,744	$ 54,175,000	$1,961,500,000	$ 55,520,000
1939	2,052,137,012	2,052,137,012	77,629,654	61,561,000	1,998,000,000	57,240,000
1940	2,238,118,011	2,025,508,207*	76,615,971	60,765,000	2,[illegible]	62,460,000
1941	2,973,485,949	2,699,083,059*	91,764,084	80,575,000	2,765,000,000	82,920,000
1942	4,280,397,059	3,802,476,135*	118,745,535	106,740,000	3,[illegible]	110,450,000
1943	5,718,447,552	5,095,045,257*	165,105,898	152,341,000	5,318,000,000	115,540,000
1944 (est.)	6,100,000,000	5,400,000,000*	170,347,631	162,000,000	5,[illegible]	150,150,000
1945 (est.)	4,207,000,000	3,840,000,000*	124,800,000	115,250,000	5,068,000,000	112,050,000
1946 (est.)	3,633,000,000	3,270,000,000*	106,275,000	98,100,000	3,732,000,000	101,370,000
1947 (est.)	3,767,000,000	3,390,000,000*	110,175,000	101,700,000	3,505,000,000	105,050,000

* Does not includes wages in excess of $3,000 per year from any single employer.

† Includes all wages in employment covered by the Act without regard to the limit of $3,000 per year per worker for tax purposes.

‡ Actual total contribution rate, because of experience rating, was 3.15 per cent instead of 3.7 per cent (2.7% from employer and 1% from employee).

§ Assumes distribution of earnings in excess of $3,000 per year to be that indicated in Table 2, and Total Taxable Wages during these years to equal (approximately) 90 per cent of Total Salaries and Wages, a proportion which maintained in 1940.

SOURCE: Actual figures compiled by the State of California, Department of Employment, Division of Research and Statistics from E.S.-203 Reports. Estimates made under assumption that (1) War in Europe will be terminated by July, 1945; (2) War in Orient will be terminated by July, 1946; and (3) Ratios which have prevailed between total Wages and Salaries in Covered Employment and those subject to the California Unemployment Insurance Act during 1940–1944 will, in general, continue through the period 1945–1947.

ity to share in the benefits of the program, these data are important in indicating the number of workers who will qualify under the plan as having earned the minimum amount of wages or more during the year.

The next step in developing actuarial estimates is to determine the amount of taxable wages upon which, under the proposed provisions, a

3 per cent payroll tax would be levied for the health insurance fund. These estimates are presented below in Table 3. Considerable light on the situation, it will be observed, has been shed by the records of the California Employment Stabilization Commission, which since 1938, has been collecting a payroll tax on all taxable wages and salaries in covered industries. By projecting the same relationships which the taxable wages (subject to the California Unemployment Insurance Act) have borne to these total salaries and wages (Table 1) and total wages in covered employment (Table 3), it is possible to estimate the probable revenues flowing into a health insurance fund year by year for the ten-year period 1938–1947. Under the proposed 3 per cent compulsory health insurance payroll tax, a fund of probably more than $100,000,000 a year for the next three years could be anticipated for the purchase of medical care under the program.

II

Estimates of the Number of Employees and Dependents in California Who Would be Brought Within the Scope of the Governor's Proposal

Obviously the number of workers included in any health insurance measure will depend upon the income levels established and the extent and nature of specific exemptions that are made. The most reliable estimates of the number of wage earners throughout the state receiving earnings of varying amounts have been made by the Employment Stabilization Commission, and are set forth above in Table 2. As previously noted, many thousands of wage earners working in covered employment in California make unemployment insurance tax contributions during the course of the year even though their total earnings during the year do not reach $300, the amount stipulated under the Act as being the minimum earnings which qualify a worker to receive benefits under the program.

The figures presented in Table 4 set forth, upon a basis of the records of the California Department of Employment, the actual monthly average of the total number of workers in covered employment coming within the scope of the Unemployment Insurance Act of California, and the estimated number of workers in covered industries with wage credits in excess of $300 (and who are, therefore, eligible for the receipt of unemployment compensation benefits under the California Act) for the years 1938–44 inclusive; and estimates for the years 1945–47.

It is important to note that the estimated number of workers eligible for unemployment compensation benefits, as set forth in Table 4, is approximately 20 per cent higher than the actual monthly average of those who are at work in covered employment. Thus it will be seen, for instance, that in 1940 approximately 1,661,000 workers were eligible

TABLE 4

NUMBER OF RESIDENT CIVILIANS, NUMBER IN LABOR FORCE, NUMBER GAINFULLY EMPLOYED, AVERAGE MONTHLY NUMBER IN COVERED EMPLOYMENT (U. INS. ACT) IN CALIFORNIA

1938–1947

Year	Resident Civilian Population	Estimated % of Gainfully Employed to Total Pop.	Number Gainfully Employed (Civilians)‡	Number in Covered (U. Ins. Act) Employment (Monthly Average)§: Number	Number in Covered (U. Ins. Act) Employment (Monthly Average)§: % of total population	Estimated Number Eligible for U. C. Benefits‖
1938	6,687,000	35	2,341,000	1,299,[illegible]	19.7	1,537,000
1939	6,755,000	35	2,361,000	1,282,[illegible]	18.8	[illegible]
1940	6,843,000	36	2,482,000	1,384,346	20.2	1,661,000
1941	7,000,000	39	2,730,000	1,672,154	23.9	2,006,000
1942	7,315,000	42	3,072,000	1,981,953	27.1	2,378,000
1943	7,510,000	46	3,461,000	2,274,631	30.3	2,729,000
1944	7,880,000	44	3,448,000	2,202,958	28.0	2,644,000
1945	8,004,000	43	3,352,000	2,061,000	25.9	2,491,000
1946	8,343,000	38	2,170,000	1,919,000	23.0	[illegible]
1947	8,440,000†	35	2,975,000	1,688,000	20.0	2,026,000
1950*	High 9,510,000	35	3,340,000			
	Median 8,650,000	35	3,050,000			
	Low 7,800,000	35	2,750,000			

* Estimates of the California State Reconstruction and Reemployment Commission, presented in *Maximum Estimates of Population Growth in California, 1940–1950* (June, 1944), p. 70.

† Population and Gainfully Employed estimates are based upon the average of the maximum and minimum estimates made by the SRRC. *Ibid.*, p. 31.

‡ Includes wage and salary workers, employers, own account workers, and unpaid family workers. Figures for 1943 (June) and 1944 (May) are from the SRRC *op. cit.*, Appendix, Table 11.

§ These figures of average monthly employment are not to be confused with the number of persons with wage credits in any year, or the number having wage credits in excess of $300 and so eligible for unemployment benefits. Out of a total of some 3,800,000 persons with some wage credits in 1944, approximately 2,650,000 had wage credits in excess of $300 each, but only 2,204,000 were actually employed in June 1944 in covered employment.

‖ Assuming that at any given time there will be approximately 20% more persons with wage credits in excess of $300 during the current year than are actually employed at the time.

for unemployment compensation benefits in California; in 1942 this number increased to an average of 2,378,000; and in 1943 a peak was reached, at which time some 2,729,000 workers were eligible for benefits under the Unemployment Insurance Act. By 1947 it is estimated that this total number will decrease to slightly more than 2,000,000 workers (2,026,000).

The Governor's bill specifically included as dependents only those who are legally dependent spouses or dependent children under the age of eighteen (Administration Bill #800). It is extremely difficult to arrive at a fair estimate of what these numbers will be per insured worker. Perhaps the simplest and most logical approach to this calculation, at least until more accurate estimates can be made, would be to base the computations upon the ratios that prevailed at the time of the last Federal Census (1940).

TABLE 5

COMPUTATION OF RATIO OF AVERAGE NUMBER OF PERSONS NOT GAINFULLY EMPLOYED TO AVERAGE NUMBER OF PERSONS GAINFULLY EMPLOYED IN CALIFORNIA, 1938–1947

Year	Total Civilian Population (Resident)	Total Gainfully Employed (Covered and Non-covered Employ) (Civilian)	Total No. Not Gainfully Employed	Ratio of No. Not Employed to No. Employed
1938	6,687,000	2,341,000	4,346,000	1.86
1939	6,755,000	2,364,000	4,391,000	1.86
1940	6,843,000	2,482,000	4,361,000	1.76
1941	7,000,000	2,730,000	4,270,000	1.56
1942	7,315,000	3,072,000	4,243,000	1.38
1943	7,510,000	3,464,000	4,046,000	1.17
1944	7,880,000	3,448,000	4,432,000	1.29
1945	8,004,000	3,362,000	4,642,000	1.38
1946	8,343,000	3,170,000	5,173,000	1.63
1947	8,440,000	2,975,000	5,465,000	1.84

These census data show that in 1940 out of a total civilian population of 6,843,000 a total of 2,482,000 persons were gainfully employed in both covered and non-covered employment throughout the State. This meant that there were at that time a total of 4,361,000 who were not gainfully employed. In other words, for every person gainfully employed at that time there was an average of 1.76 persons not gainfully employed. Projecting this relationship analysis through each year of the period (1938–47), we find that the ratio of the persons not gainfully employed to the total number gainfully employed diminished from 1938 steadily down to 1943, and upon a basis of our estimates for the coming years will increase again to the point whereby at 1947 it will have regained its pre-war level of approximately 1 to 1.84.

Another method of estimating the total number of dependents to be included under the Governor's proposal is to examine the relationship between the number of persons gainfully employed (in both covered and non-covered employment) and the total number of families

throughout the State. The ratios reflecting this relationship during the ten-year period under observation are presented in Table 6.

These computations have been made under the assumption that the average number of persons per family within the State (3.2) as revealed by the Federal Census of 1940 will have remained substantially unchanged during the 1938–1947 period.[1] The trends in this ratio from 1938 throughout the ten-year period are significant in the completion of the actuarial estimates herein undertaken. Beginning with a level of 1.12 wage earners per family, on the average, in 1938, the ratio reached

TABLE 6

COMPUTATION OF ESTIMATED AVERAGE NUMBER OF WAGE EARNERS PER FAMILY IN CALIFORNIA

1938–1947

Year	Resident Civilian Population*	Estimated Number Families†	Number Persons Gainfully Employed‡	Estimated Average No. Wage Earners Per Family
1938	6,657,000	2,069,[illegible]	2,341,000	1.12
1939	6,755,000	2,110,938	2,364,000	1.12
1940	6,843,000	2,138,438	2,492,000	1.18
1941	7,000,000	2,187,500	2,730,000	1.25
1942	7,315,000	2,285,938	3,072,000	1.34
1943	7,510,000	2,346,875	3,464,000	1.49
1944	7,880,000	2,462,500	3,443,000	1.40
1945	8,004,000	2,501,250	3,362,000	1.34
1946	8,343,000	2,607,188	3,170,000	1.22
1947	8,440,000	2,637,500	2,975,000	1.13

* Federal Census to 1940; from that point estimates are made upon a basis of the [illegible] [illegible].

† Computed upon basis of 3.2 persons per family in California, which was the average at the time of the 1940 Federal Census. (*Statistical Abstract of the United States*, 1942, p. 48.) The average number of persons per family in California at the time of the 1930 Federal Census was 3.34 (*op. cit.*, 1937, p. 45).

‡ Estimates as made in Table 4, except for Census year 1940.

a peak four years later at which time there were throughout the State an estimated 1.48 wage earners per family. The ratio then declines during the remaining years of the period under analysis, until 1947, at which time the average number of wage earners per family probably will have returned to the pre-war ratio of approximately 1.13 (wage earners) to 1 (family).

There appears to be a simple way of making the adjustments which have been made necessary by these far-reaching changes in the conditions of employment. In Table 7 is presented the estimated average number of dependents per employed worker. These computations show the ratio from year to year, based upon the assumption that dependents

[1] This assumption appears justified in the light of the slight change in the average size of California family which, according to the Federal Census, took place between 1930 (3.34) and 1940 (3.2). (See notes under Table 6).

include only the legal spouse and dependent children under the age of 18 years.[4]

TABLE 7

ESTIMATED AVERAGE NUMBER OF DEPENDENTS (LEGALLY DEPENDENT SPOUSE AND DEPENDENT CHILDREN) PER GAINFULLY EMPLOYED WORKER IN CALIFORNIA, 1938–1947

Year	Total Population (Resident Civilian)	Total Gainfully Employed	Aver. No. of Dependents per Employed Worker (Dep. spouse and dep. children under 18) (AB #800)
1938	6,687,000	2,341,000	1.31
1939	6,755,000	2,394,000	1.31
1940	6,843,000	2,482,000	1.24
1941	7,000,000	2,730,000	1.10
1942	7,315,000	3,072,000	.97
1943	7,510,000	3,464,000	.82
1944	7,880,000	3,448,000	.91
1945	8,004,000	3,362,000	.97
1946	8,343,000	3,170,000	1.15
1947	8,440,000	2,975,000	1.30

The 1940 Federal Census showed that out of the total resident civilian population of 6,843,000 persons in 1940, 2,482,000 were gainfully employed. Another 3,065,871 were reported as married dependents not gainfully employed and dependent children (not gainfully employed) under 18 years of age. In other words, at the time of the last Federal Census, there existed an average of 1.24 dependents (as defined in line with the provisions of Governor Warren's Bill) for each gainfully employed worker in California.

It should be noted also that the year to year ratios have been corrected for the increases brought on by wartime employment opportunities which have had the effect of (a) increasing the proportion of those gainfully employed, and (b) reducing the proportion of the number of dependents. Both of these influences have had a considerable effect upon the amount of tax contributions into the proposed health insurance fund, and the average amount per person available for medical benefits.[5]

[4] As provided for in Administration Bill #800.

[5] The year to year variations in this set of ratios have been computed by taking the ratios of dependents per employed worker determined by the 1940 Federal Census (when the relationship between the total number not employed to the total number employed was 176 to 100) as 100 per cent (base year), and then adjusting each year to a corresponding relative ratio. Thus the link relatives become 1938–106, 1939–106, 1940–100, 1941–89, 1942–78, 1943–66, 1944–73, 1945–78, 1946–93, and 1947–105. Thus the ratio of dependents per worker in 1940 (1.24) becomes the base, and the adjusted ratios for 1938 are 106 per cent of 1.24, etc. This adjustment has the effect of correcting for the unusually high employment levels during the peak war years of 1942–45 and therefore results in conservative rather than inflated estimates relating to the size (per capita) of the health insurance fund.

III

Estimates of the Average Per Capita Amounts Made Available for Medical Benefits Through the Health Insurance Fund

It now becomes possible to draw the various sets of estimates together in determining the approximate *per capita* amounts which such a plan would have available for covering the cost of extending medical benefits under the proposal. This brings the actuarial analysis to the crux of the compulsory health insurance proposals, for unless adequate funds are available, medical services cannot be offered to beneficiaries, and the system is faced with financial disaster.

TABLE 8

COMPUTATION OF ESTIMATED AVERAGE NUMBER OF PERSONS (WORKERS AND DEPENDENTS) ELIGIBLE FOR MEDICAL BENEFITS UNDER COMPULSORY HEALTH INSURANCE PROPOSALS IN CALIFORNIA, TOGETHER WITH ESTIMATED PER CAPITA AMOUNTS AVAILABLE FOR EXPENDITURES FOR MEDICAL BENEFITS*
1938–1947

Estimates Limited to Those Earning $4000 and less per year in Covered Employment				Estimated Total Tax Contributions (3% rates)	Aver. Per Capita Amts. Available for Med. Benefits
Year	Workers	Dependents	Total		
1938	1,517,000	2,013,000	3,530,000	$ 55,920,000	$15.75
1939	1,523,000	1,995,000	3,518,000	57,240,000	16.27
1940	1,661,000	2,060,000	3,721,000	62,460,000	16.79
1941	2,006,000	2,207,000	4,213,000	82,980,000	19.69
1942	2,376,000	2,307,000	4,683,000	119,430,000	25.49
1943	2,720,000	2,238,000	4,957,000	159,560,000	32.12
1944	2,614,000	2,406,000	5,050,000	170,100,000	33.70
1945	2,401,000	2,329,000	4,730,000	119,060,000	25.17
1946	2,303,000	2,648,000	4,951,000	101,370,000	20.47
1947	2,020,000	2,634,000	4,660,000	105,080,000	22.45
Ten-year aver.			4,465,000	$109,325,000	$23.46

* Assuming same industry group coverage as those groups of workers presently covered by the State Unemployment Insurance Act. Excluded therefore are Federal and State employees, and certain other groups which might be included under the Governor's proposals.

Comparisons of the estimated total number of insured persons, together with the total tax contributions in covered industries under a compulsory system, and per capita averages, year by year, throughout the ten-year period are set forth in Table 8. As must be expected in view of constant fluctuations in employment and earnings, these data indicate a marked variation from year to year in the estimated total tax contributions accruing to the Health Insurance Fund. In the face of inevitable changes in the estimated total numbers of beneficiaries,

these changes mean variations in the average amount per beneficiary available for the purchase of health services. The estimated per capita amount, for instance, which would be available under the general provisions of A.B. #800 increases from $15.75 in 1938 to $16.79 in 1940, $25.49 in 1942, and $33.70 in 1944; it then declines to $22.55 in 1947.

In a critical analysis of these basic estimates it is important to examine the averages that may be expected to maintain throughout the entire ten-year period. Especially does this appear to be advisable, since wartime conditions have already brought, and promise to continue to bring, basic changes in the economy of the State. Averages over the full period 1938–1947 tend to minimize both the highs and the lows of individual years within the period, and present steadier estimates over a longer period of time. Hence such averages might be expected to indicate with reasonable conservatism the conditions in general to be experienced during the starting years of a compulsory health insurance plan.

These ten-year averages, set forth in Table 8 above, suggest with reasonable certainty that amounts in the health insurance fund will average, year in and out for the coming few years, substantially more than $20 per annum per insured person. In fact the yearly estimates (demonstrated previously to have been made conservatively) during 1945, 1946 and 1947, fall as low as $20.47 during one year only, and average over the ten-year period more than $23.00 per capita per year.

These data indicate strongly that for the immediate years ahead there should be available within the Health Insurance Fund under the Governor's proposal some $20–$22 per beneficiary (both worker and dependent) for the extension of medical benefits as provided under the plan, and exclusive of administrative costs.

APPROXIMATION OF CHI-SQUARE BY "PROBITS" AND BY "LOGITS"

JOSEPH BERKSON, M.D.
Division of Biometry and Medical Statistics, Mayo Clinic, Rochester, Minnesota

If a function $\hat{Q}$ is fitted to a series of observed rates q, the value of chi-square for the ith rate q_i based on n_i individuals is

$$\chi^2 = \frac{(X-m)^2}{m} = n_i \frac{1}{\hat{P}_i \hat{Q}_i}(q_i - \hat{Q}_i)^2. \tag{1}$$

In the method employed by Bliss,[1] and Bliss and R. A. Fisher[2] for fitting the integral of the normal curve with the use of a linear transformation as "probits," there is implicit an approximation

$$q - \hat{Q} \doteq \hat{Z}(p_r - \hat{P}_r) \tag{2}$$

in which p_r and $\hat{P}_r$ are the probits corresponding to q and $\hat{Q}$ respectively, the probit being 5 plus x in the normal integral

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-x^2/2}dx = \text{rate}$$

$\hat{Z}$ is the ordinate of the normal curve at x for the fitted rate $\hat{Q}$ and $\hat{P} = 1-\hat{Q}$. If the approximation (2) for $(q-\hat{Q})$ is substituted in (1), we have as an approximation of chi-square

$$\chi^2 \doteq n_i \frac{\hat{Z}_i^2}{\hat{P}_i \hat{Q}_i}(p_r - \hat{P}_r)^2. \tag{3}$$

In the method of Bliss and Fisher, the quantity on the right of (3) summed for all the observations, is minimized. This is accomplished by obtaining a least squares solution for $\hat{P}_r$ which is a straight-line function of the log dose, the weights for the solution being

$$n_i \frac{\hat{Z}_i^2}{\hat{P}_i \hat{Q}_i}.$$

The logistic may be fitted with the use of logits by a method precisely analogous to that used for probits. The approximation corresponding to (2) would then be

$$q - \hat{Q} \doteq -\hat{P}\hat{Q}(l - \hat{L}) \tag{4}$$

TABLE 1

COMPARISON OF LOGIT AND PROBIT APPROXIMATIONS TO χ^2 WITH ACTUAL VALUE OF χ^2

q	Q	l logit q	L logit Q	p_r probit q	P_r probit Q	$\frac{Z^2}{PQ}$	χ^2 Actual $\frac{1}{PQ}(q-Q)^2$	χ^2 Logit $pq(l-L)^2$	χ^2 Probit $\frac{Z^2}{PQ}(p_r-P_r)^2$	Ratio to actual χ^2: Logit	Ratio to actual χ^2: Probit
.1	.105	2.197223	2.142862	3.718448	3.746435	.351840	.000266	.000266	.000275	1.00^-	1.03
.105	.1	2.142862	2.197223	3.746435	3.718448	.342218	.000278	.000278	.000268	1.00^-	0.96
.1	.2	2.197223	1.386294	3.718448	4.158379	.489866	.062500	.059185	.094808	0.95	1.52
.2	.1	1.386294	2.197223	4.158379	3.718448	.342218	.111111	.105217	.066233	0.95	0.60
.4	.5	0.405464	0.000000	4.746653	5.000000	.636620	.040000	.039456	.040861	0.99	1.02
.5	.4	0.000000	0.405464	5.000000	4.746653	.621919	.041667	.041100	.039918	0.99	0.96

in which l and $\hat{L}$ are the observed and fitted logits respectively, the logit being

$$l = \ln \frac{p}{q} \qquad \hat{L} = \ln \frac{\hat{P}}{\hat{Q}}.$$

If the approximation (4) for $(q-\hat{Q})$ is substituted in (1), we have as the approximation for chi-square

$$\chi^2 \doteq n_i\hat{P}_i\hat{Q}_i(l_i - \hat{L}_i)^2. \tag{5}$$

The quantity on the right of (5) summed for all the observations is then to be minimized. This is accomplished by obtaining a least squares solution for $\hat{L}$ which is a straight line function of the log dose, the weights for the solution being $n_i\hat{P}_i\hat{Q}_i$.

In (5) there is a simplification as compared with (3) because $\hat{Z}$, which is the first derivative of the function with respect to the linear transformation and for the probit is equal to the normal ordinate, is for the logistic equal to $\hat{P}\hat{Q}$, and so cancels out that value in the denominator of (1).

If "corrected" logits are to be used analogous to those used for probits, a similar simplification is effected. The "corrected" probit is a value p', substituted for the observed p_i, such as to fulfill (2) exactly instead of approximately. For the probit this gives

$$p_i' = \hat{P}_i + \frac{(q - \hat{Q})}{\hat{Z}}. \tag{6}$$

If analogously we evaluate a "corrected" logit l' to substitute for the observed l in order to make (4) hold exactly, we have for the "corrected" logit

$$l' = \hat{L} - \frac{(q - \hat{Q})}{\hat{P}\hat{Q}}. \tag{7}$$

In the method advanced by me in a former article for the logistic,[1] a different approximation than (4) is utilized

$$q - \hat{Q} \doteq -\sqrt{pq\hat{P}\hat{Q}}\,(l - \hat{L}). \tag{8}$$

Substituting approximation (8) in (1) gives as the approximation for chi-square the sum of which is to be minimized

$$\chi^2 \doteq n_ip_iq_i(l_i - \hat{L}_i)^2. \tag{9}$$

In the use of (3) for probits or (5) for logits, a preliminary fit has to be made to obtain values of $\hat{P}\hat{Q}$ which are required for the weights used

in the solution. If (9) is used, no preliminary fit is needed since the weights are in terms of the observed rates pq. The use of (9) therefore considerably simplifies the solution.

It is however pertinent to determine the relative precision of the approximation of chi-square (3) utilized in the probit method and (9)

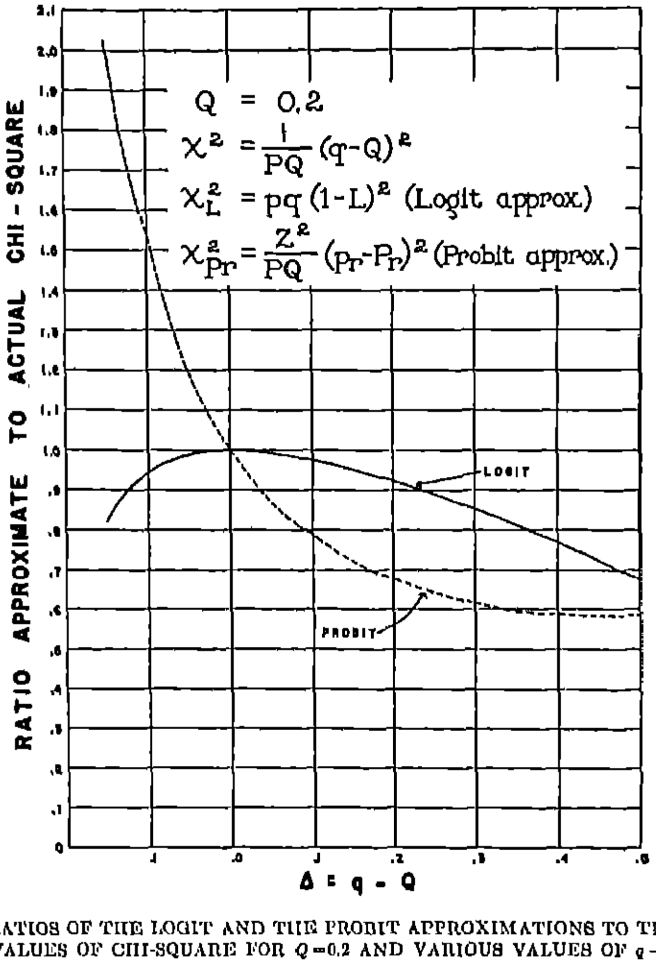


FIG. 1. RATIOS OF THE LOGIT AND THE PROBIT APPROXIMATIONS TO THE EXACT VALUES OF CHI-SQUARE FOR $Q=0.2$ AND VARIOUS VALUES OF $q-Q$.

used in the logit method. This has been attempted by taking some specific hypothetical examples and arithmetically evaluating the comparison. In Table 1 is shown for a range of q and Q, where q is considered the observed and Q the corresponding "true" rate, a comparison of each approximation with the actual chi-square. In Figure 1 is shown

the locus of the ratio of the approximate to the actual chi-square for the value $Q=0.2$, and various differences $(q-Q)$.

The following appears to be generally true: (a) The approximation of both probits and logits approaches the correct value of chi-square as the difference $(q-Q)$ approaches zero or as q or Q approaches 0.5 or as both conditions are met; (b) the approximation of logits given by (9) is closer than the approximation of probits given by (3); (c) the logit approximation gives always a smaller value than the actual, the probit sometimes lower, sometimes higher; (d) the variability of the difference from actual is greater for the probit than for the logit.

REFERENCES

[1] Bliss, C. I.: The determination of the dosage-mortality curve from small numbers. *Quart. Jour. Pharm. & Pharmacol.* 11: 192–216, 1938.

[2] Bliss, C. I. (with an appendix by Fisher, R. A.): The calculation of the dosage-mortality curve. *Ann. Applied Biol.* 22: 134–167, 1935.

[3] Berkson, Joseph: Application of the logistic function to bio-assay. *Jour. Am. Stat. Assoc.* 39: 357–365 (Sept.) 1944.

SEDLEY ANTHONY CUDMORE 1878–1945

THE DEATH of Dr. S. A. Cudmore, head of the Dominion Bureau of Statistics in Canada on October 17th while attending the United Nations Food and Agricultural Conference in Quebec City brought to a close the career of a distinguished Canadian public servant. His mastery of a wide range of economic, statistical and historical knowledge enabled him to carry out the responsibilities of Dominion Statistician with great success. He came to the Bureau shortly after its formation in 1918 and contributed much to its rapid development in the various positions he held since that time.

He was born in County Cork, Ireland, on November 27th 1878, and was a son of the late Thomas Gardiner Cudmore and his wife Caroline Ellen Sedley. His maternal grandfather was Lieutenant Colonel Anthony G. Sedley, a veteran of the Peninsular Wars and Waterloo and Governor of the Military Knights of Windsor.

Emigrating to Canada when he was nine he lived in Brampton, Ontario, where he obtained his public and secondary education. His insatiable thirst for education had to be satisfied the hard way, his high school career having been interrupted by pecuniary difficulties which led him to become a printers' apprentice at the age of thirteen. After resuming his studies he matriculated and won the Prince of Wales Scholarship for general proficiency.

In 1899 he entered the University of Toronto utilizing his training as a compositor during vacations and for shorter periods to earn the funds necessary to defray his expenses. In 1905 he graduated with high honours having won a gold medal in classics and the Flavelle Travelling Scholarship. At Wadham College, Oxford, he studied history and economics, taking his B.A. degree and later his M.A. at that University.

Inclined towards journalism he did some work for London papers and was, for a time, a sub-editor on the *London Standard*. In 1908 he returned to Canada to become instructor in Political Economy at the University of Toronto. By 1917 he had become assistant-professor.

In 1919 he gave up teaching to take the post of Chief of Education Statistics in the recently created Dominion Bureau of Statistics. From 1920–1935 he was chief of General Statistics and Editor of the Canada Year Book. Under his leadership the latter steadily gained in comprehensiveness and quality until it ranked among the best in the world.

The good reputation which the Bureau of Statistics had won in other countries led to an invitation from Palestine for someone to assist them in the organization of their statistics. Dr. Cudmore was loaned for this purpose in 1935. He remained there as Government Statistician until

1938 and organized their Bureau of Statistics. Returning to Canada in 1939 he was appointed Assistant Dominion Statistician and in 1942, on the retirement of Dr. R. H. Coats, he became Dominion Statistician.

Less than a year ago his Alma Mater recognizing his distinguished services, conferred on him the honorary degree of Doctor of Laws.

He was the author of two books: *A History of the World's Commerce* and *Applied Economics* and contributed many articles to economic journals and other publications.

His extensive statistical and economic knowledge led to his attendance as economic adviser to the Canadian Government delegation at the Imperial Conference in London in 1926. He was a member of the American Statistical Association, and served as Assistant Treasurer for Canada from 1941 until his death. He was a fellow of the Royal Economic Society, the Royal Statistical Society and the Royal Society of Canada; one of the three Canadian members of the International Institute of Statistics and a member of the Inter American Statistical Institute. He had accepted the position of Secretary and Vice-President of the Canadian Political Science Association and President of the Professional Institute of the Civil Service of Canada.

In 1910 he married Phoebe Amelia Magee. His son James served with the Royal Canadian Artillery in Europe.

HERBERT MARSHALL

PROCEEDINGS

105TH ANNUAL MEETING

STATLER HOTEL, CLEVELAND, OHIO

PROGRAM

Thursday, January 24, 1946

—9:30 A.M.—

FACTORS DETERMINING INCOME AND EMPLOYMENT
(With Econometric Society)

Chairman: Arthur Smithies, Bureau of the Budget

A Macro-Economic System, U. S. A., 1920–1941
Lawrence R. Klein, Fellow of the Social Science Research Council
Interrelationships of Income, Prices, Demand and Production at Full Employment
W. D. Evans, Jerome Cornfield, Marvin Hoffenberg, Bureau of Labor Statistics

—2:30 P.M.—

FACTORS DETERMINING INCOME AND EMPLOYMENT (*continued*)
(With Econometric Society)

Chairman: O. C. Stine, Bureau of Agricultural Economics

Some Factors Governing the Length of Inventory Cycles
Lloyd R. Metzler, Board of Governors of the Federal Reserve System
The Relationship between the Rate of Savings and National Income
Wladimir Woytinsky, Social Security Board
Capital Expansion and Full Employment
Evsey D. Domar, Board of Governors of the Federal Reserve System

LABOR FORCE MEASUREMENT AND NATIONAL EMPLOYMENT POLICY

Chairman: Howard B. Myers, Committee for Economic Development

Objectives, Uses and Types of Labor Force Data in Relation to Economic Policy
Louis J. Ducoff and Margaret Jarman Hagood, Bureau of Agricultural Economics
Recent Experience with Problems of Measurement
Gertrude Bancroft and Emmett H. Welch, Bureau of the Census
Employment Statistics in the Planning of a Full Employment Program
Charles Stewart, Bureau of Labor Statistics, and Loring Wood, Bureau of the Budget

NUMERICAL SOLUTION OF REGRESSION EQUATIONS
(With Institute of Mathematical Statistics)

Chairman: W. Edwards Deming, Bureau of the Budget

A Machine for the Determination of Correlation and Regression Coefficients
Guy Orcutt, Massachusetts Institute of Technology

A Square Root Method for the Solution of Regression Equations
D. B. Duncan, Royal Australian Air Force

Error Control in Matrix Calculation
F. E. Satterthwaite, Aetna Life Insurance Company

The Compact Computation of Canonical Correlations
P. S. Dwyer, University of Michigan

Friday, January 25, 1946

—9:30 A.M.—

CONSUMER EXPENDITURES AND SAVINGS

Chairman: Morris A. Copeland, National Bureau of Economic Research

Measuring and Forecasting Consumption
Frank Garfield, Board of Governors of the Federal Reserve System

Estimating Post War Savings
George Katona, Bureau of Agricultural Economics

Determinants of National Purchasing Power
Everett E. Hagen, Office of War Mobilization and Reconversion

Discussion: Margaret G. Reid, Bureau of Human Nutrition and Home Economics
Margaret Scattergood, American Federation of Labor

USES OF STATISTICAL ANALYSIS IN WAGE-PRICE POLICY

Chairman: Winfield Riefler, Institute for Advanced Study

Uses of Statistical Analyses in National Wage-Price Policy
Harry Magdoff, Department of Commerce

Uses of Statistical Analyses in Industry Wage-Price Policy (Illustration: automobile and steel)
Harold H. Wein, Department of Justice

Uses of Statistical Analyses by Labor in Wage-Price Policy (Illustration: General Motors)
Louis H. Bean, Bureau of the Budget

BIOLOGICAL ASSAY
(Biometrics Section of the ASA)

Chairman: E. J. deBeer, Wellcome Research Laboratory

Experimental Design for Slope-Ratio Assays
C. I. Bliss, Connecticut Agricultural Experiment Station

Statistical Aspects of the Measurement of Virus Activity
W. C. Price, University of Pittsburgh

Clinical Tests on Comparative Effectiveness of Analgesic Drugs
E. M. Jellinek, Yale University

An Analysis of Collaborative Chick Assays for Vitamin D
Paul Homeyer, Iowa State College

ESTIMATING RELATIONS FROM NONEXPERIMENTAL OBSERVATIONS
(With Econometric Society and Institute of Mathematical Statistics)

Chairman: Mordecai Ezekiel, Department of Agriculture

The Economist's Problem of Statistical Inference
Jacob Marschak, Cowles Commission
Prediction and Structural Estimation
Leonid Hurwicz, Cowles Commission

—12:00 M.—

LUNCHEON—CONFERENCE ON CENSUS TRACTS

Chairman: Howard Whipple Green, American Statistical Association Census Tract Committee

Some Problems Incident to the 1950 Census Tract Program
Howard G. Brunsman, Bureau of the Census
Problems of Tracting Metropolitan Districts
C. E. Batschelet, Bureau of the Census
Summary of the Questionnaires from 60 Key Persons
Esther M. Wright, Bureau of the Census
How Census Tracts Have Been Put to Work in Cleveland (Illustrated)
Howard Whipple Green, Secretary, Cleveland Health Council and Director, Real Property Inventory of Metropolitan Cleveland
Discussion: Census Tract Key Persons and Others

LUNCHEON—BIOMETRICS SECTION BUSINESS MEETING

Chairman: C. I. Bliss, Chairman of the Biometrics Section of the American Statistical Association

—2:30 P.M.—

MEASUREMENT OF PRODUCTIVITY IN THE UNITED STATES

Chairman: Everett Hagen, Board of Governors of the Federal Reserve System

W. D. Evans, Bureau of Labor Statistics
Charles E. Young, Westinghouse Corporation
Discussion: Martin R. Gainsbrugh, National Industrial Conference Board

ESTIMATING RELATIONS FROM NONEXPERIMENTAL OBSERVATIONS (*continued*)
(With Econometric Society and Institute of Mathematical Statistics)

Chairman: R. L. Anderson, North Carolina State College
Iterative Computation Methods in Estimating Simultaneous Relations
Tjalling Koopmans and Roy B. Leipnik, Cowles Commission
Multivariate Analysis in Economics
Gerhard Tintner, Iowa State College

Experimental Designs and Their Analysis

(Biometrics Section of the ASA with the Institute of Mathematical Statistics)

Chairman: Gertrude M. Cox, North Carolina State College

On the Use of Orthogonal Functions in the Analysis of Incomplete Latin Squares
D. B. DeLury, Virginia Polytechnic Institute
Use of Adjusting Factors in the Analysis of Data with Disproportionate Subclass Numbers
R. E. Patterson, Texas A. & M. College
Selection of Sample Size for Detecting Treatment Differences
A. M. Mood, Iowa State College
Rectangular Lattices
Boyd Harshbarger, Virginia Agricultural Experiment Station

Saturday, January 26, 1940

—9:30 a.m.—

An Appraisal of Industry and Business Statistics

Chairman: Winfield Riefler, Institute for Advanced Study

Introductory Statement:
Willard L. Thorp, Department of State
Chemical Statistics
C. W. Bell, Hercules Powder Company
Market Analysis Statistics
D. R. G. Cowan, Cleveland
Retailing Statistics
E. V. Gault, University of Michigan
Electrical Goods Statistics
Stanley Holme, General Electric Company
Development of an International Statistical Program for Pulp Producers
Oliver M. Porter, U. S. Pulp Producers' Association
Food Statistics
F. E. Ridder, General Foods Corporation
General Summary:
John W. Love, Scripps-Howard Newspapers

Sampling in the Social Sciences

(With Econometric Society and Institute of Mathematical Statistics)

Chairman: W. Edwards Deming, Bureau of the Budget

Problems and Methods of a Sample Survey of Business
M. H. Hansen, Bureau of the Census
Problems of Area Sampling in Agriculture
J. R. Goodman, Bureau of the Census; and
E. E. Houseman, Bureau of Agricultural Economics
Problems of Area Sampling in Population Surveys
B. J. Tepping and J. Steinberg, Bureau of the Census

STATISTICAL PROBLEMS OF PERSONNEL MANAGEMENT

Chairman: Dwight A. Stewart, Radio Corporation of America

Statistical Needs in Personnel Administration
W. D. Evans, Bureau of Labor Statistics
Statistical Problems of Job Evaluation
Franklin G. Moore, Western Reserve University
Statistical Problems in Worker Evaluation
D. J. Bolanovich, Radio Corporation of America
Prepared Discussion: C. H. Lawshe, Purdue University

—12:30 P.M.—

LUNCHEON—STATISTICAL NEEDS FOR ECONOMIC POLICY FORMATION

Chairman: Leonard P. Ayres, Cleveland Trust Company
Theodore O. Yntema, Committee for Economic Development

Prepared Discussions: Andrew Court, General Motors Corporation
A. Loveday, Princeton University
A. Smithies, Bureau of the Budget
K. Williams, Board of Governors of the Federal Reserve System

—2:30 P.M.—

SAMPLING IN THE SOCIAL SCIENCES (*continued*)

(With Econometric Society and Institute of Mathematical Statistics)

Chairman: Samuel S. Wilks, Princeton University

The Problem of Non-response
W. N. Hurwitz, Bureau of the Census
Relative Accuracies of Systematic and Stratified Random Sampling for a Specified Class of Populations
W. G. Cochran, Iowa State College
*On the Design of a Sample of Dealers' Inventories**
W. Edwards Deming, Bureau of the Budget; and
Willard Simmons, Office of Price Administration

HOW CAN INDUSTRY AND BUSINESS STATISTICS BE EFFICIENTLY COLLECTED?

Chairman: Winfield Riefler, Institute for Advanced Study

(A panel discussion by the speakers from the two sessions "An Appraisal of Industry and Business Statistics" and "Statistical Needs for Economic Policy Formation.")

—3:30 P.M.—

THE U. S. GOVERNMENT PROGRAM FOR STATISTICS

Chairman: William R. Leonard, Bureau of the Budget

Earl Constantine, National Association of Hosiery Manufacturers

* See page 10 of this issue.

legion users of dial telephones are indebted to his mastery of the theory of probability and the practical use of statistics for prompt connections at minimum cost whenever they use dial phones.

Hugo Muench

Rockefeller Foundation. Graduate of Cornell, Washington, and Johns Hopkins Universities; doctor of medicine and public health. He has used statistics extensively in connection with international health problems such as the control of yellow fever.

E. Grosvenor Plowman

Vice President United States Steel Corporation. Graduate of Dartmouth, Harvard, Denver and the University of Chicago; at present a director of the Association; formerly a teacher and now a business administrator; leader in the use of statistics to make business function more effectively.

Leslie E. Simon

Colonel United States Army, Director of the Ballistic Research Laboratory of the Aberdeen Proving Ground. Graduate of West Point. Recognizing early the importance of quality control in the manufacture of munitions for an industrial war, he has been a pioneer in this area of the application of statistics.

Mary Van Kleeck

Russell Sage Foundation, New York City. Graduate of Smith College; former vice president of the Association; and for many years Chairman of its Committee on Employment Statistics; internationally known for her researches in social problems in which the use of statistics has played an important part.

Abraham Wald

Professor of Mathematical Statistics, Columbia University. Graduate of the University of Vienna. A research worker who has made important contributions to the development of pure mathematics as well as economic statistics and especially to the mathematical theory of statistics.

Frederick V. Waugh

Office of the Director of War Mobilization and Reconversion in Washington. Graduate of Massachusetts Agricultural College, Rutgers, and Columbia; teacher and specialist in agricultural economics and statistics. Chief of the Marketing Research Division of the Bureau of Agricultural Economics and staff member of the War Food Administration. He has used statistics and statistical methods extensively in the administration of agricultural programs.

Abel Wolman

Sanitary engineer. A graduate of Johns Hopkins University; teacher and

consultant; internationally known as an authority on engineering who has made dry statistics fluid in the solution of problems on water supply without which, regardless of the belief of some, cities cannot survive.

THEODORE H. BROWN, *Chairman*
LOWELL J. REED
JOSEPH S. DAVIS
ARYNESS J. WICKENS
HAROLD HOTELLING

Walter A. Shewhart, President of the Association for 1945, delivered his Presidential Address,[1] *The Advancing Statistical Front*, after which the meeting proceeded to the business of the evening, with Dr. Shewhart presiding.*

The Secretary, Lester S. Kellogg, gave the Annual Report of the Board of Directors.[2]

Aryness Joy Wickens, author of the amendment to Article III of the Constitution, which was passed at the 104th Annual Meeting, rose to withdraw her amendment which provided that "The number of Fellows shall not exceed three hundred at any time." Mrs. Wickens stated that after a year's work with the Committee on Fellows, she felt the amendment to be undesirable. Chester I. Bliss, who had seconded her amendment, agreed to its withdrawal. However, since the amendment had been approved at the previous meeting, the By-Laws required that it be put to a vote. A vote was taken and the amendment was defeated.

The amendment to Article XII of the Constitution, which was passed at the 104th Annual Meeting, was presented to the membership for its approval and was passed unanimously.

Article XII—Amendments

The Constitution may be amended either at an Annual Meeting of the Association or by mail ballot. A proposed amendment must be submitted by at least ten members of the Association in writing to the Board of Directors which in turn shall submit it with its recommendation to the members of the Association not less than 60 days prior to the Annual Meeting at which it is to be voted upon or not less than 60 days prior to the date upon which the mail ballot is to be taken. The adoption of a proposed amendment shall require the affirmative vote of two-thirds of the members present and voting at the Annual Meeting at which it is voted upon, or if the amendment is *voted upon by mail ballot, of two-thirds of the members whose mail ballots* are returned within 30 days after the date of original mailing.

The Secretary recommended that By-Law number 2 which provides that "Regular Members and Fellows shall pay annual dues of $5.00 payable at the beginning of the calendar year, . . ." be changed to read: "Regular Members and Fellows shall pay annual dues of $5.00, $3.00 of which is

* The Board of Directors having received no correction to the minutes of the 104th Annual Business Meeting, they were considered approved as printed in the March 1945 JOURNAL.

[1] Printed on page one of this issue.

[2] Printed on page 88 of this issue.

for one year's subscription to the *Journal of the American Statistical Association* and 50 cents for one year's subscription to the *American Statistical Association Bulletin*, payable at the beginning of the calendar year, . . ." The amendment was passed.

W. G. Cochran reported for the Resolutions Committee (W. G. Cochran and E. Grosvenor Plowman) and proposed the following resolutions:

1. Whereas, the National Bureau of Economic Research has completed its first quarter century of service to the Nation, and

Whereas, under the leadership of Dr. Wesley C. Mitchell, this Bureau has made an outstanding contribution in its numerous research projects and activities,

Therefore, be it resolved, that the American Statistical Association, in annual meeting assembled, hereby approves this richly deserved resolution of congratulation for work well done, and especially commends the high standards of statistical technology evidenced in the publications of the bureau.

And, be it further resolved that the American Statistical Association felicitates its former President, Dr. Wesley C. Mitchell, for his great work as Director of Research of the National Bureau of Economic Research.

2. Resolved that the officers and members of the American Statistical Association express their sincere appreciation of the devoted services rendered under most trying circumstances during the past year and their gratitude to its Secretary-Treasurer, Lester S. Kellogg, for his invaluable contribution to the development and welfare of the Association.

3. Resolved that the officers and members of the American Statistical Association express their gratitude to the Secretary of Labor and the Acting Commissioner of Labor Statistics for their contribution to the development and efficacy of statistical science by sharing with the Association the services of Lester S. Kellogg.

4. Whereas, despite difficult travel and other conditions, the Annual Meeting of the American Statistical Association has been successful from every standpoint,

Therefore, be it resolved, that the special thanks of the Association are hereby extended to the members of the Cleveland Committee on Local Arrangements, particularly to Mr. E. A. Stephen, Mr. John W. Love, Miss Rose Vormelker, as well as to their associates and to the Cleveland Convention Bureau for their generous and unflagging assistance and cooperation, without which this Annual Meeting could not have been held.

5. Whereas, the Program Committee has assembled an outstanding group of speakers at short notice,

Therefore, be it resolved, that the Association takes this means of thanking the members of the Program Committee for their effective contribution to the success of this meeting.

6. Whereas, during 1945, the American Statistical Association has made further substantial progress in all its activities,

Therefore, be it resolved, that the Association hereby records its praise and thanks for the unselfish and successful efforts of its officers, its extremely capable staff, its chapter officers and committee members.

The resolutions were passed unanimously.

Helen M. Walker, Chairman of the Nominating Committee, reported that the following persons were nominated for officers of the American Statistical Association for 1946:

President: ISADOR LUBIN, Commissioner of Labor Statistics
Vice-Presidents: F. L. CARMICHAEL, University of Denver
DOROTHY SWAINE THOMAS, Giannini Hall, University of California
S. S. WILKS, Princeton University
Directors: (For the terms expiring at the Annual Meeting in 1948)
WALTER A. SHEWHART, Bell Telephone Laboratories
SAMUEL A. STOUFFER, Army Service Forces
Secretary-Treasurer: LESTER S. KELLOGG

The Secretary reported that he had received no nominations by petition and the slate of the Nominating Committee was unanimously elected.

Isador Lubin, newly elected President of the Association took the Chair of the meeting and Willard L. Thorp, Chairman of the Committee on Structure, reported on the draft constitution prepared by his committee.[1]

Rufus S. Tucker, proposed the following Amendment to the By-Laws of the Constitution:

> "There shall be a Committee on Professional Standards, composed of five Fellows appointed by the Board of Directors, whose duty shall be to recommend to the Board of Directors such actions and procedures as may tend to maintain and enhance the reputation of the Association and of the statistical profession for veracity and accuracy."

Dr. Lubin suggested referring the amendment to the Board of Directors for action and Dr. Tucker accepted the proposal.

The meeting was adjourned.

LESTER S. KELLOGG, *Secretary*

[1] See report of Committee on Structure printed in this issue.

Report of the Board of Directors

The year 1945 has been another year of steady advance on a wide statistical front.

With the coming of peace, we will gradually be able to travel again, to have meetings of the whole Association and its committees, and to engage more actively in the publication programs which are so essential to the growth and development of the profession of statistics. The signs of this development have been important in the activities of the American Statistical Association in 1945.

Meetings. Soon after the defeat of Japan, the Board of Directors and Officers decided to plan an annual meeting of the Association at Cleveland, Ohio between January 24 and 27, 1946, at the same time as the annual meetings of the Institute of Mathematical Statistics, the American Economic Association, the Econometric Society, and the American Marketing Association. The members of the Program Committee, the Officers and members of the Cleveland Chapter, the District Representative, and many members of the Association immediately directed major attention to the development of a program and to the arrangements of physical detail for the meeting. The annual meeting is the first of the whole Association's membership since December 1941—the first in a break of more than 4 years. At this same meeting, the Board authorized the appointment of a committee to prepare programs for meetings to be presented at the sessions of the AAAS at St. Louis, Missouri, March 27, 28, 29, 1946.

It is the sincere hope of the Board of Directors that there will never be another catastrophe of such calamitous proportions as to stop major activities of the nation's professional groups.

Operating under a Constitution which did not provide for such restrictions as the war imposed, the Board of Directors found it impossible during the war to carry on certain kinds of Association activities satisfactorily. At an Annual Business Meeting of the membership, called by the Board of Directors at the time of the regional meetings in Washington, D. C. between December 27 and 29, 1944, a proposal for revision of the Constitution was passed which will make it possible for the membership to vote on issues submitted by the Board in the interim periods between annual meetings.

The Board of Directors, as in other years, has met regularly, at times for extended sessions of as long as 2 days to face the various and new problems of interest to the statistical profession.

Chapters. The Chapters of the Association have had valuable programs of meetings during the year. Reports of these meetings, which have been published in the ASA *Bulletin*, have indicated that the Chapters are sharply conscious of the problems facing the development of data, techniques, interpretation and professional stature. A new Chapter was chartered in Seattle, Washington, bringing the total number of Chapters to 24 at the year's end. Inquiries from several other centers of population indicate that additional petitions for Chapter charters may be expected in 1946.

Publications. Following the resignation of Dr. Frank Ross as Editor of the JOURNAL, the Board of Directors established a new Editorial Board consisting of a Managing Editor of all Association publications, a Technical Editor of the JOURNAL, a Review Editor, and six other members. Twelve Editorial Collaborators to assist the members of the Editorial Board, both in stimulating the writing of papers and in refereeing and reviewing articles, have also been appointed. The names of members working on these groups are listed at the beginning of this JOURNAL. The Editorial Board has held two meetings during the year to lay plans for the Board's activities and procedures. Editorial policy was reviewed and a restatement of it has been proposed which will soon be presented in the ASA *Bulletin;* a restatement of the JOURNAL's book review policy was presented in the ASA *Bulletin* in August, 1945.

Plans are being made to devote one issue of the JOURNAL in 1946 to "The Role of Statistics in the War."

At its meeting on December 29, 1944, the Board of Directors approved a budget for 1945 which made it possible to introduce the *Biometrics Bulletin*, publication of the Biometrics Section. The *Biometrics Bulletin* is edited by the Biometrics Section Editorial Board, and with the assistance of the Secretary's office in production has successfully completed the first year of six bimonthly issues. Through loyal support of the membership of the American Statistical Association and several membership campaigns, the total distribution of the *Biometrics Bulletin* numbered 1214 as of December 31, 1945.

The Membership Directory, giving more detail about members and more classifications than any previous American Statistical Association Directory, was issued with the December 1945 issue of the JOURNAL. Plans are being made to issue a supplement containing names and addresses of all new members in each December issue of the JOURNAL.

The ASA *Bulletin* has continued to be issued four times a year.

Committees. The war continued during 1945 to take a very serious toll in terms of restrictions on committee activity. Heavy individual wartime responsibilities and restricted travel and hotel facilities made a wide range of committee activity impossible. In spite of these difficulties, however, there has been very important committee activity and substantial contributions therefrom to statistics.

The following standing committees have continued activity and with the return to peace expect opportunity for greater activity in 1946: Committee on Census Enumeration Areas, The Census Advisory Committee, and the Committee on Occupational Classification.

The Committee on Fellows took the Directory personnel forms as a basis and have devised a new procedure for reviewing the accomplishments of members who should be considered for election as Fellows. The committee has also completed a report which was published in the November issue of the ASA *Bulletin*.

The special committee which was appointed in 1943 by President Goldenweiser at the request of Secretary of Labor Perkins was discharged with the thanks of the Board for the excellent job it had done.

The Association's Committee on Education and Training of Statisticians has progressed less far than was hoped, due to the resignation of its chairman.

Upon special assignment, a committee working with a similar group from the Institute of Mathematical Statistics drafted a "Description of the Profession of Statistics," for use of the National Roster of Scientific and Specialized Personnel of the War Manpower Commission, as a guide for returning servicemen. Parts of this statement have been widely used in statements published by the Army and Navy.

The special Committee on Structural Organization early in 1945 invited representatives of a number of professional societies to meet to discuss the possibility of revising the Constitution of the American Statistical Association, to make it a society that could more effectively represent the interests of statistics in all substantive fields The representatives of the other societies agreed that the proposal was good and asked the Association representatives to proceed with a draft Constitution and to present the draft to a subsequent meeting of the group. The draft was presented to the group late in the year. Suggestions for revision were received and incorporated in the draft Constitution which was presented to the annual meeting and will be distributed to the membership in 1946.

A special Committee was appointed to work jointly with a similar committee of the American Economic Association on the preparation of a Source Book of Economic Statistics. The committee was composed of the following: Walter Mitchell, Jr., J. Frederic Dewhurst, Philip M. Hauser, and Theodore O. Yntema.

Problems of the utilization and disposition of government war records have attracted a great deal of attention in the latter part of 1945. A special committee composed of Lester S. Kellogg, Morris A. Copeland, William R. Leonard, and Aryness Joy Wickens was appointed to work on these problems and to represent the American Statistical Association in similar efforts which may be made by other professional associations.

At the request of Secretary of Labor Frances Perkins, an Advisory Committee to the Bureau of Labor Statistics to review the Bureau's projected employment statistics program was appointed. This committee has worked diligently and soon will present a detailed report to the Labor Department. The committee was composed of: S. S. Wilks, W. A. Berridge, Robert Burgess, O. A. Fried, and M. I. Gershenson.

Employment. During 1945 the Secretary's office has continued to provide a skeleton type of placement service for statisticians. Plans for expansion of such service are still in a discussion stage.

HENRY B. ARTHUR	E. GROSVENOR PLOWMAN
CHESTER I. BLISS	WALTER A. SHEWHART
LESTER S. KELLOGG	WILLARD L. THORP
SIMON KUZNETS	HELEN M. WALKER

Report of the Secretary

The Secretary's activities during 1945, as in all other years, are closely related to every phase of the Association's program. Since the beginning of the annual publication of a summary of the work of the Board of Directors, in 1938, the Secretary's report has dealt mainly with membership changes.

During 1945, membership in the American Statistical Association increased by 444, bringing the total as of December 31 to 3,480. This growth during the last of the war years if of great importance as an indication of the potential interest in the Association's development. It is fitting for the Secretary at this time to provide more information than is customary on the Association's membership over an extended period.

Listed below are the figures on total membership at the end of specified years. The complete story is not told by these figures because there have been several periods of decline, especially during the depression years of 1933 to 1935, and again in 1941 and 1942:

1914:	700
1918:	800
1935:	1,750
1939:	2,500
1945:	3,480

The 1945 total is composed of the following classes of members:

Honorary Members	15
Fellows*	123
Regular Members	3,342
Total Membership	3,480
Corporate Members	4

One hundred sixty five members were still in the armed forces of the United States at the end of the year. There were 61 resignations and 100 members were removed from the active rolls for failure to pay dues.

Donald R. Belcher, Alfred Cahen, Erwin A. Gaumnitz, David H. McAlpin, Kenneth A. Norton, Millard H. Pryor, Winfield W. Riefler and Elmo Roper were elected to life membership during 1945, bringing the total Life members to 48.

The death of the following members was recorded during the year: Edwin F. Gay and. E. W. Kemmerer, *Fellows and Senior Members;* David Friday, *Fellow,* Richard T. Ely and J. H. Hollander, *Senior Members;* William W. Adams, C. M. Barnes, Walter W. Cooper, Sedley A. Cudmore, Lawrence Flinn, Lyonel C. Florant, Harry L. Gale, Dorothy F. Holland, George E. Mason, Norman Merriam, John F. Normano, E. F. Perkins, John B. Shephard, Lewis R. Soffer, Arthur E. Thompson, Edith Wood, and Leo Day Woodworth, *Regular Members*, and Delphine P. Harriman, *Associate Member.*

LESTER S. KELLOGG, *Secretary*

* This figure does not include the twelve fellows elected at the Annual Meeting in January 1946.

Report of the Treasurer

The year 1945 has followed the general pattern of 1944 and 1943. The budget has had to be kept continuously fluid.

The 1945 budget as approved by the Board of Directors planned for an income of $21,385 and expenditures of $25,230. As the year ended, income totaled $24,430 and expenditures $25,098, reflecting a deficit about one-quarter as large as that planned. The major reasons for the increase of actual income over planned income was the success of a continuous membership drive, of increased subscriptions, and the income from the *Biometrics Bulletin* which had not been estimated.

Larger expenditures than those planned were due mainly to increased costs of printing the JOURNAL and the greater than estimated cost of printing the *Biometrics Bulletin*. The increase in cost of the JOURNAL was due primarily to additional expenses involved in shifting the editorship. Greater than anticipated expenses for the *Biometrics Bulletin* were due to the large printing of early editions in order to obtain wide initial distribution. The effects of efforts expended on membership and subscription drives were reflected in additional office expenses, which were considerably higher than planned.

At the current rate of expansion of activity and growth in membership, it is likely that the 1946 budget of income and expenses will see another rise over the 1945 budget and it is hoped that this budget will likewise have to be kept continuously flexible and expansible.

LESTER S. KELLOGG, *Treasurer*

Report of the Auditors

To the Board of Directors of
American Statistical Association

We have examined the attached financial statements of American Statistical Association relating to the year ending December 31, 1945. Our examination was made in accordance with generally accepted auditing standards applicable in the circumstances, and included such tests of the accounting records and other supporting evidence and such other procedures as we considered necessary.

The recorded cash receipts for the year were traced to the deposits shown on the bank statements and the amounts for dues and subscriptions were tested with the membership and subscription records. The paid checks and relative vouchers were inspected in support of the cash disbursements for the year. The bank balances were reconciled with certificates obtained direct from the depositaries and the cash on hand and the securities owned at December 31, 1945 were verified by inspection. We did not check the membership and subscription records in detail or make any independent verification of the inventory of old JOURNALS, the office records of which are based in part on data assembled in prior years, no recent physical inventory having been taken.

The life membership reserve is computed on the basis of the combined annuity table of mortality with assumed interest at 4% per annum and an assumed annuity of $5.00 per life member, in accordance with a resolution of the Board of Directors on March 31, 1930. The amount treated as income in each year represents the excess of the reserve at the beginning of the year plus interest for the year and new life membership receipts over the required reserve at the end of the year.

In our opinion, the accompanying statements present fairly the position of American Statistical Association at December 31, 1945 and the results of its operations for the year in conformity with generally accepted accounting principles applied on a basis consistent with that of the preceding year.

PRICE, WATERHOUSE & CO.

Washington, D. C.
March 6, 1946

AMERICAN STATISTICAL ASSOCIATION
BALANCE SHEET

Assets	December 31, 1945	December 31, 1944
Cash in bank and on hand	$11,383.58	$10,178.04
Accounts receivable	80.04	289.53
Investments:		
United States Savings Bonds, Series D, at redemption value	5,749.00	5,615.00
Stocks, at cost, (at market quotations $6,875 and $5,100, respectively)	5,793.50	5,793.50
Inventory of old JOURNALS, at approximate cost	1,807.56	1,683.32
Furniture and equipment, at cost less depreciation	828.20	716.52
	$25,641.88	$24,275.91
Liabilities		
Accounts payable	$ 1,772.07	$ 1,247.16
Deferred income (collections applicable to subsequent year):		
Dues	$ 6,570.00	$ 5,295.72
Subscriptions	3,064.85	2,711.57
JOURNAL advertising	183.76	176.40
	$ 9,818.61	$ 8,183.69
Net worth:		
Life membership reserve	$ 3,201.75	$ 2,614.83
Centenary Sustaining Fund, per statement	6,255.25	6,114.21
Surplus, per statement	4,594.20	6,116.02
	$14,051.20	$14,845.06
	$25,641.88	$24,275.91

AMERICAN STATISTICAL ASSOCIATION
INCOME STATEMENT

	Year ending December 31, 1945	Year ending December 31, 1944
INCOME:		
Dues—current year	$15,410.50	$12,002.75
Dues—prior years	101.75	47.75
Life memberships	317.07	230.08
Subscriptions	4,588.01	3,000.00
Advertising	897.80	940.50
Reprints	470.23	974.70
Journal sales	947.35	1,207.35
Index to the JOURNAL, sales		10.00
Special publications		56.33
Biometrics Section income	1,587.83	
Annual meeting		322.86
Miscellaneous	103.16	12.50
Dividends and interest (after deducting $104.59 in 1945 apportioned to life membership reserve)	277.11	170.72
	$24,420.37	$20,871.72
EXPENSES:		
Journal—printing, mailing and reprints	$ 6,074.02	$ 5,008.63
Bulletin	717.02	672.51
Salaries and wages	12,157.14	11,480.02
Unemployment compensation tax	11.37	10.00
Rent	1,020.00	834.00
Office supplies, printing and mimeographing	1,285.03	680.66
Postage	871.30	678.65
Telephone and telegraph	247.44	200.41
Travel expense—officers	270.72	523.11
Biometrics Section expenses	2,570.40	91.03
Directory expense		85.00
Storage of old JOURNALS	72.00	72.00
Cost of old JOURNALS sold	92.10	154.88
Depreciation of furniture and equipment	146.15	120.44
Miscellaneous	420.40	361.00
	$25,942.19	$21,145.80
Balance, loss, carried to surplus	$(1,521.82)	$ (274.08)

AMERICAN STATISTICAL ASSOCIATION

SURPLUS STATEMENT

	Year ending December 31, 1945	1944
Balance at beginning of year	$6,116.02	$6,400.10
Deduct—Excess of expense over income for the year, per income statement	$1,521.82	$ 274.08
Appropriation to Centenary Sustaining Fund of amount equal to net proceeds from sales of Index to the JOURNAL		10.00
	$1,521.82	$ 284.08
Balance at end of year	$4,594.20	$6,116.02

CENTENARY SUSTAINING FUND

	1945	1944
Balance at beginning of year	$6,114.21	$5,670.81
Add—Amounts received from contributors on pledges	$ 30.00	$ 327.50
Interest received on bank savings account and increase in redemption value of securities	106.54	105.90
Net proceeds from sales of Index to the JOURNAL	4.50	10.00
	$ 141.04	$ 443.40
Balance at end of year	$6,255.25	$6,114.21

Committee on Structure

Although the Committee on Structure unfortunately reached a high coefficient of scatteration during the last year, it has substantial progress to report. At its request, the Board of Directors issued invitations to American Marketing Association, American Public Health Association (Vital Statistics Section), American Standards Association, Econometric Society, Institute of Mathematical Statistics, National Quality Control Society, New York Society of Security Analysts, Population Association, Psychometric Society, and the Sociometric Institute, Inc. to attend a conference to discuss the overall problem of organization in the statistical field. For some years, there has been a tendency for specialized groups within the general field of statistics to set up their own specialized organizations. In spite of the substantial amount of common interest, there has been little relationship among these groups, other than from interlocking memberships. It is obvious that this tendency was in substantial degree the result of the failure of the American Statistical Association to meet the needs of these specialized groups.

The conference came to the conclusion that the time has come to bring these various organized activities in the statistical field into a closer relationship and to provide machinery whereby hereafter the satisfaction of specialized needs would be given substantial encouragement. Some sort of central body seemed necessary to act as spokesman for all statisticians, to eliminate unnecessary duplications in costs and energy exhaustion, to integrate and cross-fertilize meetings, programs and publications, and to develop organized activity in certain areas not presently occupied. It was felt that the American Statistical Association as the oldest and largest organization might well take the lead and develop a program which would be such that these other organizations might work more closely with it while not losing their autonomy in directing their own affairs.

After considerable study, the Committee has come to the conclusion that this aim can best be accomplished through a substantial reorganization of the American Statistical Association. In connection with this, provision should be made for the recognition of associate societies and affiliated societies. An associate society would be one concerned with statistical method or the statistical aspects of some specific subject-matter field. An affiliated society would be one interested in statistics but not as a matter of primary concern. Both associate and affiliated societies would be given substantial participation in the government of the American Statistical Association, and the members of associate societies would receive a regular bulletin designed for statisticians in all fields.

The reorganization of the American Statistical Association involves its internal structure as well as its foreign relations. Proposals are that there be a responsible governing body known as the Council, which will consist of representatives of associated and affiliated societies, members elected by regional groups of chapters, the Board of Directors and the officers—perhaps thirty individuals. There would be provision providing for referenda, but in general the Council would be the legislative body. The President and Board members

would be elected by the Council. The Committee has also considered at some length the problem of chapter requirements, the necessity for increasing dues, the desirability of designating a President-elect, etc., etc.

By the end of the year, the Committee had completed a draft of a revised Constitution and By-laws. It is hoped that these can be discussed during the spring at various chapter meetings, and that action on the proposals can be taken early enough to permit the new program, if and as approved, to go into effect in 1947.

WILLARD L. THORP

The Committee on Fellows.[4] The Board of Directors prior to January 1, 1945 requested the 1945 Committee on Fellows to review the whole problem of Association honors. Because the question of honors in the last analysis must be based upon the objectives sought in conferring these honors and because these, in turn, are dependent upon the objectives sought in the new Constitution, the work presented in this report can be considered only suggestive. It is recommended, consequently, that the report presented herewith, if found acceptable, shall be passed on to the Committee on the Constitution so that plans may be adopted which will be consistent with the objectives of the new Constitution.

There are at least four reasons in favor of honors which provide objectives for a plan. These reasons are:

1. A gesture of appreciation or of commendation which is official in character and granted in recognition of work of a high quality well done.
2. A device to stimulate interest in the work of the Association so that members honored will feel themselves to be more definitely an integral part of the Association.
3. An obligation to the general public since in the growing importance of statistical work it is desirable that the public look toward some group for leadership. Honors provide a means of securing such leadership.
4. A device for promoting international good will through a definite honor conferred upon a citizen of a foreign country.

It should be noted that success in these objectives, or in others which may be included, is dependent upon a well organized, aggressive Association. Such an organization should accept the responsibilities which its position implies toward problems which have now come to be of national importance in a postwar society.

In the material which follows, the recommendations are first stated. The presentation of recommendations is followed by reasons for making them.

Recommendations

I. Fellows

It is recommended that

a. Election to class of membership of the Association known as Fellows

[4] Mr. Henry B. Arthur was appointed by President Walter A. Shewhart to the Committee on Fellows on December 11, 1945 to serve for the term January 1946 through December 1950.

shall represent a distinguished honor for work of high quality in statistics.

b. Fellowship shall be limited to citizens of the United States and Canada.
c. The method of their election shall be as provided in the present Constitution.
d. Their number shall be relatively small and limited to 200 members instead of the present limit of 150.
e. The Fellows be used in various special capacities on behalf of the Association and be more directly integrated with its activities.

II. Honorary Members

It is recommended that

a. This class of membership be limited to citizens of foreign countries.
b. Their election shall be accomplished through a Nominating Committee of three Fellows who shall serve for a term of three years. Upon nomination by the Committee on Honorary Members, final election shall be by the Board of Directors, and notification made by the President of the Association in collaboration with the Chairman of the Committee.
c. The number shall not be greater than 25.

III. Special Prizes

It is recommended that

a. American Statistical Association Awards be established for members who have not reached their 32nd birthday on January 1 of a given year.
b. The Awards are to be made at the Annual Meeting.
c. Three Awards are to be conferred for not more than three outstanding papers published during the two years preceding each annual meeting. One of these papers is to be in statistical theory, one is to be an application of statistical methods and the third may be either of these types or a combination of them. No awards need to be granted if papers of sufficient quality have not been published.
d. The Awards consist of a prize of $100 and a Certificate.
e. A member shall not be eligible for a second Award.
f. No Fellow is eligible for an Award.
g. The Fellows should be polled annually for lists of distinguished papers.
h. Awards should be made by a committee of three Fellows on Awards. This committee should be organized in the same way as the Committee on Honorary Members.

IV. Distinguished Lecturers

It is recommended that

a. There be chosen each year an annual Lecturer who should address the Association at the time of the Annual Meeting, and other lecturers who, in the course of a year, should speak before the chapters.
b. The Annual Lecturer and the Chapter Lecturers may be selected from any group of the Association membership.

c. The Lectures together with the Presidential Address shall be published by the Association early in the year following each Annual Meeting.

d. The Chapter Lecturers, in general, shall be selected from sections of the country at some distance from the Chapter at which the lectures are to be given. To facilitate this work, funds should be provided to cover travel expenses of the Lecturers.

e. A committee of the Fellows appointed by the Board shall arrange for these Lectureships. The Secretary of the Association shall be a member of this Committee.

V. Finances

It is recommended that efforts be made to finance these recommendations through special funds.

Discussion

I. Fellows

The basic problem of the Fellows is two-fold: (1) whether to represent a group selected to confer an honor for accomplishment, or (2) whether they are to be selected merely as a professional class of the Association. Regardless of the group which they are to represent there will always be differences of opinion in regard to the criteria used and the wisdom of the Committee's selection. Nevertheless, it is believed on the basis of past experience that the selection can be made with a reasonable degree of precision. If the Fellowship is to represent a distinguished honor in terms of relatively long accomplishment, clearly brilliant younger members are likely to be passed by temporarily. With this in mind, the proposal of special prizes and Distinguished Lectureships has been suggested as a means of recognition by the Association for a larger group. There is the further condition that the older active members of the Association are likely to be heavily loaded with important work. If the Association can command their interest for work in perpetuating their own number and in acting as members of committees in connection with other honors, their importance in the work of the Association will be definitely enhanced.

As judged by the replies to the long questionnaire sent to the Fellows over a year ago, there is not the slightest doubt about the very real interest which many Fellows have in the Association. It might very well happen, consequently, that they could be polled from time to time in regard to other important problems or policies of the Association. Thus they would act as an informal advisory Senate for the Association officers.

There is no objective line of reasoning which will lead uniquely to the limiting number for Fellows. A review of the early 1945 Association membership, however, indicates the following facts: If the criterion of two or more check marks in the right-hand column of the card shown on page 45 of the November, 1945, *Bulletin* be used, it will be found that including Fellows there will be listed some 600 or 700 individuals. This is about 25% of the membership of the Association. A limit of 250 would represent about a third

of these. In other words, about a third of the members who have secured recognition either in Association activities or in publication of articles would be the limiting number for the honor of Fellowship. Whether this number would be desirable in the light of more critical evidence to be obtained in the years ahead is unknown. Whenever a condition of nonbalance occurs, however, the Constitution could again be amended.

The present method of determining the Committee on Fellows is wise and should be continued. It leaves the Committee completely free from influence of current officers. It provides the maximum of independence and so lends quality as well as dignity in the conferring of this honor.

II. Honorary Members

A poll of the Fellows early in 1945 showed almost unanimous support for the decision that the honorary members should be citizens of foreign countries exclusively. There was a difference of opinion as to how the election should be carried through. Nevertheless, the weight of the thoughtful comments was in the direction of the difficulty of identifying outstanding statisticians in foreign countries and, consequently, of the belief that a continuing committee organized like the Committee on Fellows and reporting to the Board of Directors would be the most effective way of accomplishing the result. Here is another extremely difficult but important problem for the Fellows to undertake.

III. Special Prizes

Much earnest comment has been received from the Fellows in regard to the necessity of honoring scholarly work of the younger members of the Association. Thoughtful comments tended to imply that the Fellows will, as a rule, be elected from among the older members of the Association who, through years of effort, had built for themselves an outstanding reputation. There was the need, consequently, for some means to confer honor upon the younger members of the Association and, in so doing, to stimulate their interest in professional work and in the Association. It was with these thoughts in mind, that the recommendations for the special prizes were prepared. Emphasis should be placed upon the intent to secure a breadth of interest in the papers and upon the use of a two-year period. This time interval obviously will result in a one-year overlap. Intention is to provide for possible recognition of both of two papers which happen to have been published in a given year on, for example, The Theory of Statistics. One paper might be recognized one year and the other at the end of the following. Another reason is to avoid excluding authors who happen to have articles appearing in December journals. The cash award is large enough to be of importance to most young men and the Certificate probably would be cherished as the first of a possible series of honors to be won during a lifetime of work.

The ineligibility of Fellows is deliberate since the intent of the prize is to discover new talent. It would be hoped, of course, that prize winners in some future year would be made Fellows.

IV. Distinguished Lecturers

Several objectives are present in this recommendation. The object of the Distinguished Lecturer at the Annual Meeting is to bring recognition to members of outstanding ability commonly in the age group between 30 and 45 who have not been elected Fellows. The intent is that the Association will be honored in the act of honoring them. On the other hand, intent of the Chapter Lectureships is to stimulate the work of the chapters and at the same time to confer directly an honor upon any member of the Association who may be chosen to deliver such a lecture. It will be noted that no special cash award of any kind aside from the distinction of Distinguished Lecturer has been included. The suggestion that transportation expense be provided is done with the deliberate hope that Lecturers can be interchanged between Chapters as widely separated as New York and Los Angeles, or Washington, D. C., and Seattle.

THEODORE H. BROWN, *Chairman*
LOWELL J. REED
JOSEPH S. DAVIS
ARYNESS JOY WICKENS
HAROLD HOTELLING

The Biometrics Section. Following approval by the Board of Directors of the American Statistical Association, the main project of the Biometrics Section in 1945 has been the *Biometrics Bulletin*. The Editorial Committee for the Bulletin consists of G. M. Cox, Chairman, C. I. Bliss, W. G. Cochran, F. R. Immer, H. W. Norton, L. J. Reed, G. W. Snedecor and Sewall Wright. Complimentary copies of one or both of the first two numbers and an invitation to join the Section were sent to all members of the parent Association and to other potential members. The response has been encouraging. As of December 31, 1945, 1079 members and associate members and 135 subscribers were receiving the Bulletin.

The *Biometrics Bulletin* has appeared bi-monthly, beginning in February. The first volume, which totals 92 pages, would have required nearly twice as many pages in the format of the *Journal of the American Statistical Association*. In order that the volume would coincide with the calendar year, the Bulletin started without a backlog of material. The Section is indebted primarily to our very able Editor for filling the Bulletin despite many difficulties, and to the Secretary's office of the Association for the production and distribution phases. As the Bulletin becomes better known and meetings are resumed, we can look forward to a steady growth both in its influence and size.

Membership in the Section on December 31, 1945 comprised 862 members and 217 associate members, the former belonging to the parent Association and paying annual dues of one dollar, the latter two dollars. Fields of interest represented in the Section have been tabulated from data on 769

individuals. Of these, 65.8 percent reported an interest in some branch of biology with medical science the most frequent, 16.1 percent an interest in meteorology, chemistry, mathematics or industry, and 18.1 percent in the social sciences. Expenses of the first year not covered by dues and subscriptions have been absorbed by the parent Association or contributed by members of the Editorial Committee and by the North Carolina State College. The parent Association has been compensated in part by new members who have joined both the Section and the parent Association. We should look forward, however, to the necessity of increasing dues in 1947.

The last sessions of the Biometrics Section were held at the annual meeting of the Association in December 1944, none having been held in 1945 due to the limitations on travel. In 1946, two sessions are scheduled for the Cleveland meetings in January including one sponsored jointly with the Institute of Mathematical Statistics. A joint session in the week of March 11 at Atlantic City is being arranged with the American Society for Pharmacology and Experimental Therapeutics. Other sessions are being planned for the meetings of the A.A.A.S. in St. Louis during the week of March 25 by a special committee under the chairmanship of H. C. Fryer. Two of these are deferred meetings from 1945.

The Biometrics Section was represented at Inter-Society conferences on February 10 and December 8 for coordinating the activities of professional organizations interested in statistical method. These plans provide for the growth of sections into self-supporting associate societies integrated into an over-all program. In line with this development, a special committee of which A. E. Brandt is chairman will report a draft constitution to the business meeting of the Section in Cleveland. If this constitution is adopted, the Section will become the "Biometrics Society, an Associate of the American Statistical Association."

CHESTER I. BLISS, *Chairman*

Committee on the Employment Statistics Program of the Bureau of Labor Statistics. This Committee was established by the Association in May, 1945, at the request of the Secretary of Labor, to review the proposed expanded employment statistics program of the Bureau of Labor Statistics. The mandate of the Committee is limited to the Department of Labor and to current statistics as distinguished from conjectural forecasts. During the summer and fall, individual members of the Committee have had numerous conferences and discussions with officials and technical experts of the Bureau of Labor Statistics. A large number of documents pertaining to procedures for obtaining estimates of employment in an expanded program have been received from the Bureau of Labor Statistics and are being studied by the Committee.

Members of the Committee have also discussed sources of information, procedures, and other matters relating to the problem of estimating employment with members of the Committee on Labor Supply, Employment

and Unemployment Statistics and with individuals in various bureaus and agencies of the government. Reports and other documents from these agencies are under study by the Committee.

A meeting of the Committee, together with representatives of the Bureau of Labor Statistics, Bureau of the Census, and Bureau of Economic Security, was held in New York on September 11, 1945, at which time various questions and topics on procedures being used or contemplated in the enlarged employment statistics program of the Bureau of Labor Statistics were discussed.

It is anticipated that a preliminary draft of a report now under preparation will be completed within a few weeks.

W. A. BERRIDGE
R. W. BURGESS
O. A. FRIED
M. I. GERSHENSON
S. S. WILKS, *Chairman*

Committee on Census Enumeration Areas. The Committee on Census Enumeration Areas held its annual meeting at a luncheon session on Friday, January 25, 1946, at the Carter Hotel in Cleveland, Ohio.

Speakers at this session were Esther M. Wright, Howard G. Brunsman, Clarence E. Batschelet, and Howard Whipple Green. The papers included a review of current activities of the tracted cities in promoting census tract use, plans of the Bureau of the Census for 1950, practical considerations in extending tracted areas beyond city limits, and techniques utilized in popularizing census tract use in Cleveland.

The meeting adjourned to the office of the Chairman of the Committee where there was an exhibit of census tract street indexes and maps, published studies incorporating census tract data, and numerous maps and charts showing the distribution of social and economic characteristics of the population by tracts. These materials included the published works from other cities as well as Cleveland. A practical demonstration of preparing spot maps and overlay maps was also given.

The meeting and exhibit were enthusiastically attended by nearly 100 persons.

During the year a summary of the questionnaires returned by the Key Persons was completed and published, 20 pages. This report includes a list of Key Persons and Census Tract Committee members for each city, facts concerning census tract maps and census tract street indexes for each city, and a list of the most important users of tract data.

A complete set of the papers presented at the 1944 December meeting, 21 pages, was also distributed to all Key Persons and others interested in census tract use.

The Committee is planning to publish a manual on census tract use during 1946.

A third of the Census Tract Key Persons left their cities during the past year but the Chairman is pleased to report that replacements have been found for all but five cities. There is definite evidence of more active interest developing currently in census tract use. The Bureau of the Census reports that its supply of 1940 tract publications has been exhausted. There has been considerable correspondence with Key Persons concerning the development of adequate census tract maps and street indexes, establishment of census tract committees, and extension of tracts in suburban areas. Several conferences on these problems have also been held with Key Persons.

The Committee wishes to express its appreciation to the Bureau of the Census for its cooperation during the year. Members of the Consulting Staff of the Vital Statistics Division have made surveys of census tract use in 20 cities as well as giving the Chairman assistance in his Committee work.

Respectfully submitted,
HOWARD WHIPPLE GREEN, *Chairman*
C. E. BATSCHELET
DOCTOR W. THURBER FALES
DOCTOR ERNEST M. FISHER
MRS. SHIRLEY K. HART
DOCTOR PHILIP M. HAUSER
DOCTOR VERGIL D. REED
DOCTOR LEON E. TRUESDELL
DOCTOR LENT D. UPSON

Joint Committee on Occupational Classification. During the past year, the Joint Committee on Occupational Classification of the American Statistical Association and the Division of Statistical Standards of the Bureau of the Budget and its technical subcommittee have been concerned with taking account of the effects of the war on occupational classification systems. This has involved the incorporation of new job titles or the reclassification of old titles necessitated as a result of war developments. Another specific problem on which considerable progress has recently been made is the translation of military classifications into their civilian equivalents. This work has facilitated the counselling and placement of veterans and of workers displaced from war industries as well as providing the basis for a better description of the occupational skills of the population in the post-war period.

The Bureau of the Census has initiated work on revising its Index of Occupations and the United States Employment Service has started a revision of its Occupational Directory. Both of these activities are long-term projects and all major classification differences arising in these two systems will be cleared through the Committee for settlement, as they have been in the past. This clearance is a continuing process and constitutes a review of the Convertibility List, which was developed by the Joint Committee on Occupational Classification in 1940 as the framework for standardization of occupational statistics.

In addition, requests for reclassification of occupations are received from the representatives of particular occupational or professional groups in the public and are considered by the Committee, with or without formal hearings. During the past year, agencies representing optometrists, chiropracters, and Christian Science practitioners have requested reclassification in the United States Employment Service and/or Census classification systems and the Convertibility List.

GLADYS L. PALMER, *Chairman*

Census Advisory Committee. The Census Advisory Committee of the American Statistical Association held only two meetings during the past year, one in April and one in November. However, from the point of view of content, the level at which the discussions were held and the problems discussed, these meetings were highly important. The appointment of a new Secretary of Commerce called for revaluation and restatement of the program of the Bureau of the Census which was reviewed and discussed to some length at the April meeting, covering not only the problems which the Bureau met in wartime but also the directions toward which the Bureau should point in the postwar transition period and in peace time. Some of these directions crystallized later in the 1947 budget request of the Bureau of the Census for the fiscal year 1947, and in order to cover this program remedial legislation was introduced into Congress. The bill which was later introduced as H. R. 4781 incorporated the suggestions which the Committee has been making over the past years and which it specifically urged in a resolution adopted in the November meeting.

In general, the program toward which the Bureau of the Census is pointing calls for benchmark censuses on a staggered basis with Agriculture and Population being taken simultaneously every five years in years ending in 0 and 5 and Manufacture and Business, also quinquennially, in years ending in 2 and 7, in addition a strong program of current data necessary for the study of economic and social data which are planned. The availability of such data would be very helpful in meeting the statistical needs of the nation. In addition to, the consideration of over-all program, topics discussed included the Census of Agriculture which was taken in the year 1945 and the organization of a Field Division in the Bureau of the Census. The latter was also the object of resolution at the November meeting, in which the Committee emphasized that the quality of statistical information as published can be no better than that obtained at the source. The existence of an adequate field staff experienced in field problems and methods would not only result in improvement in the current statistics programs now being developed but would also provide an essential nucleus around which the field staff for a major census could be organized. This would remedy many of the problems which exist by virtue of the temporary assembling and disbanding of key field personnel for the major censuses as has been the practice in the past. In addition, a continuing field organization can be effectively used on a serv-

ice basis for the collection of data for other Federal agencies. The concentration of field activity would result in a substantial improvement in Federal statistics in general and the Committee urged further study and action on this aspect of the problem.

The texts of the two resolutions follow:

Resolution on Census Bureau Program Adopted by the Census Advisory Committee of the American Statistical Association
November 10, 1945

The Census Advisory Committee of the American Statistical Association, having discussed with officials of the Bureau of the Census and with the membership of the American Statistical Association over a period of years major problems which face the Bureau of the Census, strongly urges the following conclusions and recommendations directed toward a balanced statistical program designed to be of maximum usefulness to the business community, to the government, and to the other consumers of statistical information. Such a program is urgently needed during the critical months and years ahead to effect an orderly transition from war to peace and to provide a sound factual basis for postwar adjustment.

(a) That in the coming year the Bureau of the Census be empowered to take Censuses of Manufactures and Business, a survey of the characteristics of the population on a sample basis, and a consumer income survey, and that it also be empowered to expand the present report of the labor force to obtain data for the major industrial areas.

(b) That the Bureau of the Census reschedule its basic census operations so as to achieve a better balanced program by conducting quinquennial censuses in lieu of the present unbalanced programs ranging from biennial to decennial censuses; and that these censuses be so spaced as to prevent undue administrative burden from being placed on the Bureau of the Census.

(c) That the Bureau of the Census be authorized to conduct its proposed program of current statistics including annual sample surveys and periodic monthly or quarterly reports necessary to provide timely information for current use.

It is the judgment of the Committee that both the immediate and long-run programs proposed by the Bureau of the Census would meet urgent needs of business, government, and other users of statistical information and that they should be authorized at an early date.

Resolution on Census Field Organization Adopted by the Census Advisory Committee of the American Statistical Association
November 10, 1945

One of the critical points in the collection of statistical information is the field interview, or the contact of the representative of the collection agent with the respondent. The quality of the statistical information as published can be no better than that obtained at the time of the enumeration. For this reason, the Census Advisory Committee of the American Statistical Association considers it extremely important that an adequate permanent field staff be maintained by the Bureau of the Census. Such a staff would become intimately acquainted with the problems and pitfalls of this important phase of statistical activity, would be useful in the collection of accurate data for the monthly report on the labor force, for annual surveys of retail trade and for other current sample studies. Such an organization would also provide an essential nucleus around which the field organization for the major censuses can be organized.

For these reasons, the Census Advisory Committee commends the organization of a separate Field Division within the Bureau of the Census to specialize in this activity. However, we believe that a specialized organization of this type can become most effective if it is also used on a service basis

for the collection of data for other Federal Agencies as well. This concentration of field activity in one specialized agency would result in a substantial improvement in the quality of other Federal statistics collected in the field, and the Committee strongly urges further study and action toward this end.

WILLIAM F. OGBURN, *Chairman*

List of Committees and Representatives for 1943

Committee on Fellows

- Theodore H. Brown, *Chairman*
- Lowell J. Reed
- Joseph S. Davis
- Aryness Joy Wickens
- Harold Hotelling

Committee on Nominations

- Helen M. Walker, *Chairman*
- Frederick C. Mills
- Stuart A. Rice

Committee on Investments

- Henry B. Arthur, *Chairman*
- E. Grosvenor Plowman
- Willard L. Thorp

Budget Committee of the Board of Directors

- Henry B. Arthur, *Chairman*
- E. Grosvenor Plowman
- Willard L. Thorp

Annual Meeting Program Committee

- W. G. Cochran, *Chairman*
- A. D. H. Kaplan
- Lowell J. Reed

Biometrics Section Committee

- C. I. Bliss, *Chairman*
- Horace W. Norton, *Secretary*
- Alva E. Brandt
- William G. Cochran
- Gertrude M. Cox
- Churchill Eisenhart
- Alfred J. Lotka
- Hugo Muench, Jr.
- J. Neyman
- Lowell J. Reed
- George W. Snedecor

Census Advisory Committee

- W. F. Ogburn, *Chairman*
- Murray R. Benedict
- Donald R. G. Cowan
- J. Frederic Dewhurst
- Frederick F. Stephan
- Willard L. Thorp

Committee on Census Enumeration Areas

- Howard W. Green, *Chairman*
- Clarence E. Batschelet
- W. Thurber Fales
- Ernest M. Fisher
- Shirley K. Hart
- Philip M. Hauser
- Vergil D. Reed
- Leon E. Truesdell
- Lent D. Upson

Joint Co..mmittee on Occupational Classification

Gladys L. Palmer, *Chairman* and *Secretary*
Meredith B. Givens
Ida C. Merriam
Robert J. Myers
Caroll L. Shartle
Leon E. Truesdell

Committee on Sampling Surveys

S. S. Wilks, *Acting Chairman*
W. Edwards Deming
George W. Snedecor
Frederick F. Stephan
Alfred N. Watson
Theodore O. Yntema

Advisory Committee to the Bureau of Labor Statistics

S. S. Wilks, *Chairman*
W. A. Berridge
Robert Burgess
O. A. Fried
M. I. Gershenson

Committee on the Preservation and Destruction of War Records

Lester S. Kellogg
Morris A. Copeland
William R. Leonard
Aryness Joy Wickens

Committee on Preparation of a Source Book of Economic Statistics

Walter Mitchell, Jr., *Chairman*
J. Frederic Dewhurst
Philip M. Hauser
Theodore O. Yntema

Committee on Structure

Willard L. Thorp, *Chairman*
C. I. Bliss
W. G. Cochran
F. Leslie Hayford
Frederick F. Stephan
Helen M. Walker

Representative on the Board of Directors of the National Bureau of Economic Research

Frederick C. Mills

Members of the Social Science Research Council

Harold Hotelling
Frederick F. Stephan
Willard L. Thorp

Representative on the Council of the American Association for the Advancement of Science

Walter A. Shewhart

Representatives on the Joint Committee for the Development of Statistical Applications in Engineering and Manufacturing

Walter A. Shewhart
Churchill Eisenhart

Representative on the Sectional Committee on Standards for Graphic Presentation of the American Standards Association Sponsored by the American Society of Mechanical Engineers

Frederick E. Croxton

Representative on the Advisory Board of the American Year Book

Frank W. Notestein

Representative on the Advisory Committee to the Census Library Project

Richard O. Lang

Member of the American Documentation Institute

Lester S. Kellogg

BOOK REVIEWS

Edited by
OSCAR KRISEN BUROS
Rutgers University

Industrial Quality Control. Published bimonthly by The Society of Quality Control Engineers in cooperation with The University of Buffalo. $6.00 per year; $1.25 per issue; Vol. I (112 pp., $8\frac{1}{2}\times11$ in.) is available at $4.00 per set of 6 copies or $1.00 per copy. Edited by *Martin A. Brumbaugh* (Professor of Statistics, University of Buffalo). Editorial and publication office: Crosby Hall, University of Buffalo, Buffalo 14, New York. *Two reviews follow:*

REVIEW BY SEBASTIAN B. LITTAUER
Professor of Mathematics and Chairman of the Department
Newark College of Engineering

THE appearance in July, 1944, of the first issue of *Industrial Quality Control* served notice that "Quality Control" had come of age as a field of professional endeavor. As is well known, industrial quality control in its present sense dates from the early twenties when Shewhart began his original investigations which pointed the way to the use of statistical method. Before 1930 it would have been difficult to round up a dozen papers on statistical quality control. And in spite of a definite increase, thereafter, in the industrial use of statistical techniques in quality control, and in the publication of papers on the subject, in but relatively few industrial plants, up until the war, was there an awareness of control charts or sampling tables.

None the less, the advantages of statistical quality control had been so soundly established that during the war a program of intensive training in these techniques was carried on under the auspices of the Office of Production and Research Development and the U. S. Office of Education. The enthusiastic reception of this program was followed at many of the training centers by the organization of local quality control societies. From this nucleus of persons interested in the applications of statistical methods to industrial quality control has come the demand for a journal to serve their professional needs. To quote from an editorial in the first issue: " . . . Other publications are available for the presentation of statistical and mathematical developments of interest to quality control operators. Likewise the engineering approach is available. But between these two there is an open area of analytical quality control in industrial use. Into this open area *Industrial Quality Control* is projected as a service to the men and women who make the decisions from day to day in quality control departments."

That such service is being appreciatively rendered is attested to by a substantial group of readers among whom are more than seven hundred paid subscribers. The appearance of this bi-monthly periodical has stimulated the activities of many of the young quality control societies by publicizing their

activities and reporting fully on clinics, training courses and the like; by summarizing articles of interest appearing currently in other journals; by providing text reviews, news and notes, and other material usually found in the pages of a professional journal, and in the present case for the first time in a journal of their own, appealing directly to quality control operators. The early issues contained a comprehensive bibliography which must have been heartily welcomed by many a reader. In these and other ways *Industrial Quality Control* has been rendering valuable service to its readers and encouraging a sense of common professional interest among quality control workers.

The contributed papers, which are the substance of any technical journal, are rather difficult to appraise. All of them are concerned primarily with the statistical aspects of quality control. A good third of the papers deal with "indoctrination" and "general plant-wise operations," another third with sampling plans and the remainder with control chart applications or incidental techniques. Most of the papers are by persons engaged in industrial practice. Only a few of the papers are by seasoned writers in the field, and certain names intimately associated with the development of statistical quality control are not yet among the contributors. While most of the papers contain points of interest, in a number of the papers the results are invalidated because of weaknesses in the statistical theory used to obtain them.

A periodical is far more than the sum of its parts, and, indeed, is a living entity which can exercise a vital influence on the progress of the professional field which it serves. Founding such a journal is a worthy and formidable undertaking, while nurturing it to a mature status requires and merits support from all interested quarters. To the founders and editors of *Industrial Quality Control* genuine thanks are due and all good wishes for the full realization of their objectives.

Review by Holbrook Working
Professor of Prices and Statistics, Stanford University

The origin and purposes of *Industrial Quality Control*, as stated in a "message from the President" and an "Editorial" in its first issue (July 1944) are well summarized in the following excerpts (sequence ours):

> More than a year ago the War Production Board through its Office of Production Research and Development launched a project to provide training in the use of analytical quality control to key personnel of plants producing war material[1]. . . . Early in this program an intensive course was given in Buffalo, September 11–18 [1943]. . . . On November 7 the graduates of this intensive course established the Society of Quality Control Engineers. . . . From this nucleus of 41 representatives of 15 individual industries, our membership has been augmented by graduates of three successive [part-time] classes in Quality Control, so that we now have an organization of 150 active members. . . . We propose through the medium of this publication to present

[1] On reconsideration, the author of this statement would doubtless want to acknowledge also the credit due the Engineering Science and Management War Training program of the U. S. Office of Education, which directly sponsored the intensive course with which the [illegible] started, and many part-time courses, also.

> to our membership, and the public, items of interest in Quality Control. . . . The content of each issue will be selected with the thought in mind that quality control operators in other areas face problems similar to those encountered in the Buffalo area. Accordingly materials appearing will, we hope, have more than local significance.

The objectives of the Buffalo society and of its publication were specified further in the second issue of the journal:

> Your officers are attempting to keep abreast of . . . developments by fostering two major programs: (1) the training of new quality control operators in courses offered at the University of Buffalo, (2) the publication of *Industrial Quality Control*. The magazine is designed to give you articles to show what other companies are doing, bibliographies to guide your reading concerning the use of control techniques in other parts of the country, tables to facilitate your work, and solutions of current problems presented by members of the society.

The first seven numbers of the journal were lithographed from well prepared typewritten copy on a page $8\frac{1}{2} \times 11$ inches. The first two comprised 16 pages each, including cover, and the remainder, 20 pages each. Volume II, Number 2, appeared in type, with 16 pages of the same size as before.

The articles in *Industrial Quality Control* fall mainly into three classes: (*a*) general discussions of the principles of quality control, and especially of statistical control; (*b*) reports on specific quality control projects; and (*c*) presentations of statistical theory and method applicable to quality control.

Generalizations have their merits, but one who has already read a little of the existing literature will not pause long over most of the articles which deal mainly in generalities. One, however, "Top Management Views Quality Control," by M. Herbert Eisenhart, President, Bausch and Lomb Optical Co., in the March 1945 issue, is an exceptionally thoughtful and interesting discussion of the broader aspects of quality control.

Among the papers that deal with specific applications of statistical control are two or three which include discussion of some broad problems of quality control, and in so doing are the more effective because they present conclusions in a setting of the conditions which validated them. Noteworthy on this account is "Growth of a Quality Control Department in a Canadian Ammunition Plant," by Douglas F. Brown, General Superintendent of the Technical Development and Control Department, Defense Industries, Ltd. (Cherrier Works), in the issue of November 1944. To go farther in the mention of individual papers on specific applications is scarcely possible here, for nearly every issue of the journal has contained one or two such papers of real merit, potentially useful to readers of diverse interests.

Papers and notes concerned with statistical theory and method in the early issues of the journal were pitched at too elementary a level to be of interest to most readers of this review, and they lack special merit, in general, even as elementary expositions. One in the November 1944 issue, unsigned, is particularly open to criticism. Dealing with the interesting problem of data showing highly skewed distributions in their original form, but approximating normality when transformed to logarithms, it suggests, reasonably

enough, the use of a chart for individual observations with an average at the geometric mean. It then determines "sigma" as one-third of the difference between the logarithm of the geometric mean and the logarithms of the extreme items in the distributions, and blandly accepts different values of "sigma" for positive and negative deviations. It misses the opportunity to advise plotting such data on paper with a logarithmic vertical scale. Evidencing a sense of the superiority of averages over individual observations for most control purposes, it remarks, "Perhaps readers of this article will elect to develop a solution of the problem of preparing control charts for averages and ranges of a nonsymmetrical distribution"—a suggestion that seems to reflect little understanding of the knowledge required for solution of the problem.

As a hasty improvisation under pressure of handling a plant problem, the method just summarized may be admired for its simplicity and ingenuity, but one may doubt whether similar problems were so prevalent and urgent that time could not well have been spared to develop a more efficient and rational solution before publishing on it. In any case, to allow "sigmas" obtained as just described to masquerade as standard deviations without apology is hardly excusable.

It was appropriate that discussions of statistical method in the early issues of *Industrial Quality Control* should be elementary, in keeping with the status of most of its subscribers as beginners in statistics. Later issues reflect recognition of substantial advance in the statistical competence of those whom the journal has undertaken to serve. The two latest issues (July and September 1945) contain useful short papers on "Use of Correlation in Quality Control" by Peter P. DiPaola and "Sequential Sampling Inspection for Attributes" by H. A. Freeman.

Worthy of special mention is a paper by Martin A. Brumbaugh in the September 1945 issue which presents a comparison of pre-war and present quality of the machine-shop work of one plant that adopted statistical control during the war. The paper takes advantage of an opportunity afforded by hasty storing of parts for a firm's pre-war product before screening inspection, when the plant was converted to war work. At the end of the war, when inspecting the stored parts, the chief inspector, who chooses to remain unidentified, seized the opportunity to make a record of the efficiency of pre-war control of quality. Professor Brumbaugh's comparison of the resulting data with similar information on current production shows strikingly that pre-war operations presented as much need for statistical control as was found in the production of war materials.

The Buffalo Society of Quality Control Engineers and the Editor of *Industrial Quality Control* are deserving of high praise for their publication. Its establishment and successful maintenance required imagination, initiative, and a great amount of hard work in the face of severe difficulties. Its editorial policy has shown a keen sense for the needs of its readers, and even its faults redound in considerable part to the credit of its sponsors, for they

must have known that the conditions of publication made defects inevitable, and have consciously braved criticism for the sake of rendering a service.

From the standpoint of service, one of the large contributions of *Industrial Quality Control* has been made through its pages of news and notes. Largely neglected in this review because of their generally ephemeral interest, these items have been a valued feature of the magazine, contributing substantially to advancement of the use of statistics in quality control through the encouragment that people get from knowing what others of like interests are doing.

Quality Control Reports, Nos. 1-12. Issued by the Quality Control Program, Carnegie Institute of Technology, for Office of Production Research and Development, War Production Board. *Holbrook Working* (Director, Quality Control Program; Professor of Prices and Statistics, Stanford University) and *Edwin G. Olds* (Associate Director, Quality Control Program; Associate Professor of Mathematics, Carnegie Institute of Technology) Washington: Office of Production Research and Development, Civilian Production Administration, 1945. Gratis. Paper.

No. 1, **Introduction of Statistical Quality Control at Illinois Testing Laboratories,** *James A. Fizzell* (Supervisor, Engineering Department, Illinois Testing Laboratories, Inc., Chicago). May 1945. Pp. 8.

No. 2, **Organizations Concerned With Statistical Quality Control.** *Holbrook Working* and *Edwin G. Olds*. June 1945. Pp. 38.

No. 3, **Statistical Quality Control at the St. Louis Division of American Stove Company.** *Lester A. Kauffman* (Director of Quality Control, American Stove Co., St. Louis, Mo.) August 1945. Pp. 22.

No. 4, **Quality Control at the Lukas-Harold Corporation.** *Carl G. Schmid* (Manager, Quality Control Department, Lukas-Harold Corp., Indianapolis, Ind.). August 1945. Pp. 10.

No. 5, **Statistical Quality Control in Radio Manufacture.** *Fred Trowbridge* (Quality Control Engineer) and *John T. Clark* (Production Control Engineer). (Sentinel Radio Corp., Evanston, Ill.). September 1945. Pp. 12.

No. 6, **An Application of Statistical Quality Control at the Hoover Ball and Bearing Company.** *Russell C. Trombly* (Supervisor, Quality Control Department, Hoover Ball and Bearing Co., Ann Arbor, Mich.). September 1945. Pp. 9.

No. 7, **An Adaptation of Statistical Quality Control at Aldens'.** *James M. Ballowe* (Manager, Inspections Department, Aldens' Chicago Mail Order Co., Chicago, Ill.). September 1945. Pp. 12.

No. 8, **An Application of Statistical Quality Control at the John Deere Tractor Company.** *E. L. Fay* (Chief Inspector, John Deere Tractor Co., Waterloo, Iowa). September 1945. Pp. 9.

No. 9, **Statistical Quality Control at Lockheed.** *James R. Crawford* (Manager, Manufacturing Reports Department, Lockheed Aircraft Corp., Burbank, Calif.) and *Preston C. Hammer* (Partner, Probity Engineers, Los Angeles, Calif.). September 1945. Pp. 17.

No. 10, **Some Aspects of Statistical Quality Control in the Steel Industry.** *Richard H. Ede* (Carnegie Illinois Steel Corp., Gary, Ind.) and *Earl W. Mahaney* (Metallurgist, Youngstown Sheet and Tube Co., Youngstown, Ohio). September 1945. Pp. 12.

No. 11, **Bibliography on Statistical Quality Control.** *Dorothy W. Goodfellow* (Editorial Assistant, OPRD Quality Control Program, Carnegie Institute of Technology). September 1945. Pp. 48.

No. 12, **Some Experiments Illustrating Principles of Quality Control.** *Irving W. Burr* (Associate Professor of Mathematics, Purdue University). September 1945. Pp. 12.

Two reviews follow:

REVIEW BY GEORGE W. BROWN
Research Engineer, R.C.A. Laboratories, Princeton, N. J.

THE 12 Quality Control Reports together constitute a testimonial to the success of the OPRD Quality Control Program in stimulating interest in statistical quality control in war industries and in guiding the organization of industrial quality control programs. The degree to which the subscribers to the OPRD Program have been repaid for their interest is really impressive, if these reports represent an approximately fair sample of the general experience of the organizations which availed themselves of the opportunities offered.

Nine of the reports describe individual quality control applications presented by quality control, inspection, and engineering executives covering a wide range in previous quality control experience, statistical training, and special fields, with the result that the reports themselves touch a correspondingly large variety of points of interest, including problems of organization, methods of introducing quality control techniques for the first time, and details of actual applications.

With respect to technical statistical details, there is little that need be said about these reports. With a few exceptions, they present standard control chart techniques applied to incoming material, process control, and final inspection. The greatest single gain, in the majority of cases presented, results, in the reviewer's opinion, from the combination of increased emphasis on quality at all levels with the continuous picture of the situation afforded by the control charts, providing a graphic goal of great psychological importance along with an intelligible description of the history of the process. Other gains which follow are more or less of the nature of improved inspection efficiency associated with sensible measurement and sampling, or of specification revision, made possible by knowledge of the statistical behavior of the process when in control.

The problem of whipping an out-of-control process into control is, of course, not subject to the routinization that is possible at the other control stages, nevertheless the obvious question which arises is to what extent can the quality control engineers take advantage of the well-known techniques in experimental design and analysis to shorten the time required to get a process under control for the first time. For example, it is not difficult to imagine short experiments which, without serious interference with production, would provide estimates of the variability associated with machines, operators, lots, etc., thereby facilitating the search for the "assignable causes" whose presence has been made known by the control chart. On the surface, at least, it appears likely that several of the authors might be ready to extend their services still more by the addition of analysis of variance techniques to their tools. The control chart itself is, after all, neither more nor less than an approximate analysis of variance in a particularly simple form.

Report No. 1 is of particular interest in that it describes the quality control experience of a small concern, employing fewer than 250 persons. In spite of the small size of the organization, statistical quality control was applied successfully both to incoming and plant inspection. In contrast, No. 9 deals with quality control on a gigantic scale, with statistical control being applied even to the number of parts inspected. No. 9 contains a description of a system of quality stamps for large scale operations; also of interest is a special statistical study involving inner diameters measured at four separate places in a single part.

For a detailed description of the process of locating and eliminating an assignable cause of variability, No. 5 is one of the better reports. There is also a nice example of a non-normal distribution, the distribution of the values of resistors purchased to a given tolerance, the non-normality resulting from the manufacturers' screening of resistors meeting a closer tolerance for other orders requiring the closer tolerance. The authors' example of humidity affecting the quality of the product raises the question whether the equipment in question was meant to operate satisfactorily under humid conditions. If so, a change in design should have been necessary; if not, the procedure adopted was satisfactory. This report also mentions the use of correlation methods and "t" tests in special studies.

No. 7 is unusual in that it concerns human efficiency alone, in a mail-order operation. It offers an excellent example of the effects of exhibiting a goal and keeping orderly data in a convenient form, as applied to order checking and office clerical work, where the problem is confined to one of management and employee efficiency, without the usual tool wear and material variability.

Part I of No. 10 is concerned primarily with correlation methods, rather than control chart applications, in preliminary analyses. It is interesting to note the author's stated intention of exploring the analysis of variance and experimental design techniques for controlled experiments. Part II is a detailed exposition (not entirely clear) of the process of setting up a sampling acceptance plan for process control at an intermediate stage, designed to reduce the amount of conditioning required on semifinished steel bars.

Nos. 3, 5, 6, and 8 are essentially conventional control chart applications, although No. 4 does contain a few very mild departures from "orthodox" American Standards Association procedures. No. 3 contains a very intelligent discussion about setting the mean well enough to stay within tolerances, given the variability, and on required variability limits for a given tolerance. No. 6 contains a control chart example where measurements seem to have been made to .001 when the tolerance was .002, the coarse measurement resulting in a very queer-looking chart. No. 8 has a good discussion of the problem of selling quality control to the organization and initiating the program. The highway analog is a rather amusing one, comparing the paved portion of a highway with the part between control limits, the shoul-

ders to the part between the control limits and the specification limits, and the ditches to the part outside the specification limits.

No. 2 is effectively a history of the OPRD Program, listing courses given, and the organizations and individuals connected with the various courses. The list of 810 organizations participating is certainly an impressive one, to this reviewer at least. This report, and the bibliography (No. 11), together make a useful supplement to any quality control library.

No. 12 exhibits some typical dice experiments which illustrate various quality control principles. Some of these experiments might well be listed as "musts" for individuals approaching quality control for the first time, to assist in acquiring a first hand feeling for the essential features of statistical variation. It is pointed out that p's of 1/6, or 1/3, say, obtainable with a single throw of a die, are too high to illustrate practical fraction-defective situations; the author then suggests using dice with more than six sides, or that have been "prepared." It is difficult to see why the author neglected the possibility of compounding probabilities with pairs, or greater numbers of throws, to get p's as low as desired. For example, with two throws, the probability of getting two aces is 1/36, the probability of getting an ace and a two is 1/18, the probability that the sum is at most three is 1/12, etc. It might also be pointed out that random number tables, such as may be found in Fisher and Yates, permit drawing samples from any preassigned statistical distribution, although for some purposes the effect may not be quite as satisfying as the equivalent dice experiment.

REVIEW BY LOUIS C. YOUNG
Quality Control Supervisor, Westinghouse Electric Corporation
Springfield, Massachusetts

REPORT No. 1 describes the quality control operations carried out in a small plant under war time conditions with attention centered chiefly upon screw machine operation. By means of control charts, Mr. Tizzell found that the lack of adequate maintenance was the chief drawback to combined production and quality; the charts themselves proved the means of convincing management that steps needed be taken in this direction. Examples of the application of statistical methods to problems of incoming inspection and final inspection are also given in brief, but with no tangible evidence of increased economy of operation. It is interesting to note, in the problems and their solutions, that engineering and inspection functions are as often at fault as production.

No. 2 contains summary lists of attendance at the ESMWT courses conducted by the OPRD Quality Control Program, through May, 1945. Listed in separate tabulations are: business organizations represented in attendance; the educational institutions represented in attendance; the locations at which the courses were presented, with the dates of the first and subsequent presentations; instructors in the courses; special lecturers; and a list

of the several quality control societies formed as a result of the courses presented.

Concerned with the control methods used in a medium sized plant during war time, No. 3 considers control of both variate characteristics and attributes. The description of the former is made somewhat confusing by reference to a chart with incorrectly calculated limits. A greater part of the report stresses control by attributes, both in machining threads and in assembly welding, with detailed accounts of the troubles encountered and the process changes found necessary.

Dealing with the control of machining of parts for a very precise war instrument, where production and quality evidently far outweighed expense, No. 4 centers upon a portion of the machining operations and explains the organizational steps found necessary to expedite the control program. Emphasis is placed, due to the precision of the work, upon control by variables; this is particularly of interest because of the small volume of production turned out on each order—a source of complaint for many industrial concerns attempting the use of statistical methods to control quality. Reference to charts is made slightly difficult by misplacement of the numbers of the figures.

No. 5 begins with a general discussion of experience with statistical applications, and includes a few detailed comments on their applicability where test error is high. A specific example follows, well worth reading, of an operation which was uncontrolled (from a statistical viewpoint) and in which the controlled variation would have exceeded the tolerance. Solution of this problem involved tracking down certain specified causes of extreme variation, and in the use of modified control limits along the line of variate-consumer-risk thinking which has become widespread in Great Britain during the war years.

Limiting his report (No. 6) to a single dimension commonly giving trouble in the manufacture of ball bearings, Mr. Trombly describes his attempts to increase the degree of control over the form grinding of races, by means of variate control charts. In this case, the charts were also used subsequently as a form of acceptance inspection in order to reduce the amount of detail inspection given ball races; lots characterized by samples outside the control limits were sent to the inspection crib for detailing.

No. 7 describes the reduction of errors made in filling customer's orders, in a mail order house, by means of departmental charts showing daily ratios of errors to orders filled. Although the author occasionally interjects a note on the similarity of this application to those of manufacturing, the improvements shown and the manner in which cooperation was enlisted by display of charts leads to the belief that employee cooperation or conscience may be better in merchandising than it is in industry. The description of the simplicity and directness of charting an error ratio, regardless of the cause or type of error, will appeal to those who have become lost in the multinominal

forest while trying to account separately for a variety of quality characteristics.

The two examples of improvement in quality which are described in No. 8, dealing with the control of machining processes in a tractor factory, represent appropriate applications of the methods of variables and attributes: the former, in the treatment of readily measurable expensive piston pins; the latter, in reducing rejects of threaded nuts made on automatic screw machines. Although the latter operation is one which is commonly encountered in industrial statistical literature, the account is notably straight forward, practical, and interesting in its detail with regard to sampling procedure.

No. 9, a report of statistical application in an aviation plant, is noteworthy in two respects: the wartime requirement that the quality of many key parts be assured is more applicable to post-war production in aircraft assembly plants than in many other war plants; and the principal example of statistical solution of a manufacturing problem is one in which the solution rests upon the statistical manipulation rather than upon customary common-sense consideration of inspection data,—a distinction which is less pronounced or lacking in the majority of current expositions on this subject. The use of a "quality audit" form in reporting departmental levels of quality and improvement, comparable to an expense and budget statement, serves to relate the technical and managerial phases of quality control in a practical and convincing manner.

No. 10, composed of two parts,—one dealing with the open hearth and pouring variables of molten steel; the other, with the rolling and skin conditions of steel blooms and billets in process,—is a study in the industrial application of correlation methods. In the first part, Mr. Ede describes several problems in steel manufacture, most of them concerned with the effect of chemical process factors upon the quality of steel in process, and explains the method by which each was solved and an improvement made in the process. Omission of the details on the trials and technical thought underlying the selection of method increases the clarity and brevity of this section, but detracts from an appreciation of the complexity of the problems. In the second part, Mr. Mahaney starts with the surface characteristics of steel in process and considers them as causative factors in the determination of final quality. A reduction in process time and expense is demonstrated by means of a sampling plan centering upon attribute inspection of work in process.

No. 11 is a bibliography divided into three sections, of which the first is devoted to articles on industrial statistics and quality control. This section is the most complete listing seen by the reviewer, and is laudable for the saving in time and effort it will afford those seeking information and examples of statistical control in specific industrial fields. The second section lists a number of technical publications issued by industrial firms and insti-

tutions, including those of the American Standards Association, the Bell System, the OPRD reports, and others. The third section lists books "suggested by men in various areas of statistical quality control as having been useful to them."

No. 12 describes some experiments with dice, paralleling the control chart accumulation and judgment of data. In this capacity, the report bridges the gap necessarily left in the American Standards Association war standards on quality control; by using a background of purely chance causes, the author delivers that conception of control needed by any novice in industrial statistics. At one point, Mr. Burr discusses the errors made possible by stratified sampling as opposed to rational sampling, and demonstrates by rearranging the same data into both types of sample. Although the error (of overestimating variance) is real, it is implicitly blamed upon the method of sampling, whereas it might be stated more validly as an error in the statistical treatment of data. The report also surveys, by means of similar experiments with dice, the field of acceptance sampling. The constant system of causes is used to show that an occasional sample may be so good as to justify reduced inspection, but that subsequent samples inspected under these conditions will soon cause reversion to increased inspection for the specific sampling plan and process average under consideration.

Symposium of Papers on Statistical Quality Control. Birmingham 3, England: Quality Control Panel, Birmingham District Production Committee, Ministry of Production (C.M.L. Building, Great Charles St.), [March 1946]. Pp. 76. Gratis. Paper.

Introduction, Quality Control in Industry, pp. 2–4. *W. A. Bennett* (Works Manager, English Needle and Fishing Tackle Co., Ltd., Studley, Warwickshire, England).

Lecture 1, Direct Measurement, pp. 5–8. *E. H. Sealy* (Management Consultant, Orwick, Orr and Partners Ltd., 7 Park Lane, London, W.1.).

Lecture 2, Number Defectives, pp. 9–21. *D. J. Desmond* (Research Chemist, Joseph Lucas Ltd., Great King St., Birmingham, England).

Lecture 3, Application and Organisation of Statistical Quality Control in a Light Engineering Plant, pp. 22–37 *D. S. McNaughton* (Chief Inspector, Hercules Cycle and Motor Co., Minor Works, Birmingham 6, England).

Lecture 4, Part 1. Practical Application of Statistical Quality Control in the Press Shop, pp. 38–39. *W. N. Baker* (Technical Assistant, I.C.I. Metals, Ltd., Witton, Birmingham, England).

Lecture 4, Part 2. The Practical Application and Organisation of Statistical Quality Control in Deep Drawing Press Shop, pp. 40–44. *H. C. Webb* (Chief Inspector, I.C.I. Metals, Ltd.).

Lecture 5, Sampling Schemes for Number Defectives Inspection, for Quality Estimation of Single Lots, pp. 45–66. *A. W. Swan* (The United Steel Companies Ltd., 17 Westbourne Road, Sheffield 10, England).

Lecture 6, Associated Statistical Techniques, pp. 67–74 *J. W. Rodgers* (Chief Metallurgist, English Needle and Tackle Co. Ltd.). *Two reviews follow:*

REVIEW BY IRVING W. BURR
Associate Professor of Mathematics, Purdue University

THIS book is a collection of papers discussing the application of statistical methods of quality control to industrial problems. The opening paper is a

good general introduction to the problems of where and how to use the methods in a plant. Mr. Sealy's paper explains clearly the basis of measurement control charts and the relationship of control limits and specifications, bringing in modified control limits. Mr. Desmond gives the typical application of fraction defective and then discusses the advantages of using a compressed or narrow tolerance and studying by chart the number of cases outside these limits. In this way there are enough "rejects" to enable the use of small samples. Both papers are well illustrated with diagrams and shop problems. Neither paper would be quite complete for the beginner, but both contain useful material.

The three application papers are excellent for presenting the kinds of problems which are met in practice, their solution and the general economic value of control chart methods. An industrial man will find many useful tips and adaptations, such as, a sampling compromise (last two pieces from the machine and three at random from the tote box) and the group control chart for automatics (McNaughton), the extent of applications (Baker), and the economic decision of when and when not to use the methods (Webb).

Mr. Swan's paper discusses the problem of designing single and double sampling inspection plans for attributes. The aim would seem to be to help one design his own sampling plan. The general outline for single sampling is (*a*) to choose the maximum number of pieces which can be called a lot, (*b*) to decide what is the maximum percentage of a lot which one can afford to inspect, (*c*) to determine the process average and decide upon the producer's risk at this level, and (*d*) to decide what consumer's risk to assume for some to-be-determined lot tolerance fraction. There will then be several choices of pairs of sample size and lot tolerance quality. The practical choice can be made to minimize the former or to use the lowest value of the latter. This is an interesting approach and may well be that desired by some operators. The paper is not, however, very clearly written and in particular it is difficult to find out what mathematical methods were used in the derivations. For example, the Poisson, binomial or hypergeometric distributions are not always indicated, the basis of Table IV is not presented, nor are there adequate references. On page 50 after Table III, there seems no particular reason why a sample for 105 and lot tolerance of 5 per cent would not better meet the author's desires.

Mr. Swan's double sampling plan involves two equal samples making up the equivalent of a single sample. This may or may not be an economical set-up. The section "Operating a Double Sampling System" contains two serious defects. The Operating Characteristic of a process is "the way in which the quality swings up and down about an average percentage of defectives." The point is that even when the quality level is stable at say 1 per cent defective, the fraction defective of lots actually sent will vary around this level in a random and well understood manner. Knowing the probability of acceptance of a lot of any given fraction and the probability of occur-

rence of such a level, then one can by using the latter as weights find an average outgoing quality for the manufacturing stream which is under control at 1 per cent. Average amount of sampling inspection is similarly available. Mr. Swan errs, however, in using as weights, not the probability of such lot quality levels, but the probability of such sample levels. Hence his averaged results are meaningless.

The second point is that it is actually unnecessary to use the weighting of probability of occurrence of each lot level, because if the process level is at 1 per cent this gives a constant probability of occurrence of a "pass" sample and of a "reject" sample a priori. The sampled part of the lot is not correlated *at all* with the remainder of the lot (if the process is stable at 1 per cent). It is as though we were to take a sample of 150 consecutive pieces from the process, make our decision to pass or reject and then to choose 850 more to complete the lot. The quality in the remainder will average 1 per cent in the long run regardless of what has happened in the first 150. Looked at this way it is a most simple matter to calculate the average outgoing quality. The method using lot quality weights looks convincing, sensible and correct, because it is. But so is the above simpler method. The two give the same results because both are correct.

Mr. Swan's paper has many interesting points for one familiar with say the Dodge and Romig tables (reviewed in this JOURNAL 40 (231): 382–384 S '45). It is scarcely safe nor clear reading, however, for one relatively unfamiliar with sampling inspection by attributes.

Mr. Rodgers primarily discusses two topics: tests of normality and analysis of variance. Neither topic is discussed fully enough to be very helpful to the practical man. Ranges are used for the analysis of variance, which is useful. It is not possible to check upon the accuracy of the suggested methods since the interpretations, tests and applications to be made are not indicated.

The book as a whole has much of interest to well repay careful reading.

REVIEW BY PAUL PEACH
Industrial Statistician, Institute of Statistics
University of North Carolina, Raleigh

DURING the war both British and American war agencies found it worth while to promote the use of statistics in industry. In the United States the War Production Board and the Office of Education organized short training courses in statistical quality control; there was very little supervision of actual factory installations. In England the Ministry of Production and the Ministry of Supply gathered together a number of qualified persons and set up consulting groups, or panels, whose services were made available to industry and who, it appears, worked very close to the production line. As part of their educational program they presented lectures on industrial applications of statistics. This booklet gives six lectures of a series of eight presented in Birmingham in June-July, 1944.

The content of each lecture is indicated by its title. As might have been expected, the treatment is somewhat uneven. The third and fourth lectures are severely practical; the other four are more theoretical, though not objectionably so. The lectures must have been effective and successful, but in printed form they will, in this reviewer's opinion, be interesting chiefly as history, not as reference or teaching material.

For illustration, consider the last sentence of Lecture 1, by Mr. Sealy, on control charts for measurements: "The lecture was then illustrated by several examples of charts taken from actual factory practice." The presentation and discussion of these charts must have been one of the highlights of the meeting; but nothing more appears in the printed version. Here is an extreme instance of the difference between writing for readers and writing for hearers. The speaker uses not merely words, but inflections, gesture, emphasis; he probably has a blackboard handy; he senses his audience reaction and modifies his presentation accordingly. The stenographic report of a lecture, whipped into shape for publication, often appears tremendously overcondensed, simply because so much of the ancillary nonverbal exposition is lost.

The present pamphlet suffers from this defect. A second objectionable feature has to do with the figures. Save for those of Webb and Baker, each lecture is illustrated with figures and charts. In Sealy's lecture the figures are dispersed through the text, but for the rest all figures are collected in a sort of appendix to the lecture; this makes for awkward reading. The awkwardness is increased for American readers by the use of w instead of R for range; the references, too, are chiefly to British sources.

The lectures by Sealy and Desmond present familiar elementary principles; those by McNaughton, Webb, and Baker are practical talks which must have been very effective when delivered and retain much of their effectiveness in print. The sixth lecture, by Rodgers, mentions briefly a few more advanced statistical techniques such as the chi-square test and the analysis of variance.

Lecture 5, by A. W. Swan, was of particular interest to this reviewer. It is a discussion of acceptance sampling for lot-by-lot inspection. Swan gives credit to Capt. A. H. R. Grimsey, Military College of Science, in connection with the single and double sampling plans here presented. These plans are based on the ratio "Maximum Allowable/Process Average," as are the Dodge-Romig tables of 1929 and a table recently devised by the writer and a colleague; the Swan-Grimsey single sample tables are not especially compact, and overlook the simplification that can be obtained by using chi-square. Swan and Grimsey also propose the use of two equal samples for double sampling, thus again anticipating this reviewer. The Swan-Grimsey double sampling plans are arrived at by trial and error, with considerable computation. No adequate tables are given. However effective the spoken lecture may have been, the written presentation is difficult to follow.

During the war the need was for quick results at any cost. There was no time

to give really fundamental instruction in industrial and engineering statistics. The highly condensed material and intensive instruction of war days will no doubt give place to the slower, more orderly and more thorough training of engineers and others in statistical science as part of their college education. The lectures of the present pamphlet served a purpose, and an important one; but the pamphlet itself, in the reviewer's opinion, will have little utility in the postwar period.

Industrial Experimentation. *K. A. Brownlee* (Directorate of Ordnance Factories, Explosives, London). Memorandum No. 18. London, W.C.2: S.N. 7 Branch, Directorate of Royal Ordnance Factories (Explosives), Ministry of Supply, [The Adelphi], 1945. Pp. 87. Paper. *Two reviews follow:*

REVIEW BY JOHN W. TUKEY
Assistant Professor of Mathematics, Princeton University

THE PURPOSE, scope, and origin of this pamphlet are clearly expressed in the Foreword (italics ours):

> The present memorandum has grown out of an appreciation of the desirability of subjecting experimental results to a critical statistical test of significance. It was evident that a convenient account of the straightforward tests of significance, written from the point of view of the individual who has to apply them in practice *without* necessarily *a full knowledge of their* theoretical *background*, was not readily available, and an attempt was therefore made to prepare one. As this work proceeded it became evident that, to apply tests of significance economically, the experiments had to be planned in appropriate forms. In its present form, therefore, the Memorandum is intended to be a guide to the planning and interpretation of experiments on the industrial scale, and it is hoped that its contents will become part of the everyday technique of those who carry out such experiments.

The author has carried out this task with clarity and precision, using chemical engineering examples. The pamphlet can be highly recommended to all interested in industrial experiment as a satisfactory manual of Fisherian statistics. It covers the significance of means, the comparison of variances, the χ^2 test, the analysis of variance, correlation, regression, and factorial experiment. It should prove particularly interesting, of course, to those interested in chemical engineering problems, but workers in other fields will find transfer a simple problem.

This is a brief manual for those without a full knowledge of the background of the tests—in short, a cookbook. It is a very good cookbook, whose recipes have been developed in agriculture and well tested in many other fields, but it is not, and could not be in so short a space, a treatise on the physical chemistry of cooking. The average cooking problem can be solved with a cookbook—very many industrial problems can be satisfactorily solved by mechanical use of these well-tested statistical procedures.

The distinction between statistical significance and practical significance is nowhere made clearly and precisely. The manner in which confidence inter-

vals are used in some discussions indicates the author's appreciation of this difference, yet there is no discussion of what to do if an immediate decision must be based on statistically insufficient data. There appears to be an undue tendency to believe all effects zero until statistically proved otherwise and to leave no place for parallel experience.

The brief discussion of control chart procedures gives little guidance in choosing between control chart and Fisherian methods in a specific instance, yet many of the examples which the author treats simply and directly would be difficult or impossible to treat by control chart methods. The control chart procedure was designed and is fitted to study processes which are or should be in control—that is, the successive values resemble a random sample. The Fisherian methods were designed and are fitted to study processes which are not in control, but where auxiliary data or the recurrence of the uncontrolled effects allows the partial separation of different effects. This distinction is fundamental.

In general, the possible use of methods based on order statistics is given insufficient consideration. In particular, the "maximum efficiency" of the variance is stated without restriction on page 9.

This is, of course, related to the general tendency to assume normality. Many small errors operating independently are said to produce normality—no mention that the effects must behave additively. The general principle is enunciated (p. 11) that departure from normality must be "very severe" before significance tests become appreciately inaccurate. This is not the case for Bartlett's test for the homogeneity of variances, which is described on page 18. The examples of pot life on pages 14, 28, and 31 do not seem to represent even nearly normal distributions. This is not surprising, since life is rarely normally distributed.

In carrying out the analysis of variance, the procedure used is to test doubtfully significant mean squares and, if they are not significant at the 5 per cent point, to then pool them into a new estimate of error mean square. The possible pitfalls in this process require much careful consideration. The estimation of contributions of different sources to variance is done in a similar way, making the flat assumption that each σ^2 is either actually zero or statistically significantly different from zero.

By requiring a minimum cell frequency of 5, the author restricts the usefulness of χ^2 to what now seems to be an excessive extent.

In discussing variance on page 25, the author states unreservedly that, if several factors contribute to a variance, then that variance is the sum of the component variances. This statement tacitly assumes that the factors are independent and that they act additively.

In discussing the two regression lines on page 43, the author states that he is concerned with predictive relations and not with structural relations. The reviewer cannot accept this as a general principle and feels that there are many problems of chemical engineering where the structural regression is important.

On page 34, the author assigns a certain component of variance to "bagging the material." While this is possible, it seems more likely to be associated with th' process of taking samples.

This pamphlet has now been criticized on several lines. But—the reviewer knows of no book to which the reader may go for a better treatment of these critical points. The faults are those of the current statistical literature, not of this pamphlet alone.

On page 17, the author makes his only direct slip. He uses a two-tailed F test in comparing two variances without stating that the 5 per cent level has now become the 10 per cent level, and so on.

REVIEW BY J. WOLFOWITZ
Institute of Statistics, University of North Carolina, Raleigh

This short book is intended as a description of the more important statistical methods suitable for use in chemical manufacture and is written "from the point of view of the individual who has to apply them in practice without necessarily a full knowledge of their theoretical background." Within the compass of eighty-seven pages it discusses a wide range of topics. After several introductory pages designed to explain statistical fundamentals comes a chapter devoted to the t-test. A brief chapter discusses estimation of the variance, both by point and interval. Chapters on the chi-square test, the analysis of variance, regression, correlation, and factorial design contain the classical material to be expected in such a brief treatment. Two pages are given to a discussion of the quality control chart for ranges and means; in the latter the control lines are located by using the estimate of the range. Four tables are reproduced from Fisher and Yates and one for control charts is reprinted from Dudding and Jennett's British Standard.

As another statistics book written as a manual the present one contains a number of examples drawn from the chemical industry, and is a handy compendium of a number of computational procedures. The author is obviously enthusiastic about the applicability of statistical procedures to the chemical industry and anxious to encourage their use.

A question that merits raising is whether this type of book ought not to be written somewhat differently. This reviewer would prefer to see more emphasis on fundamental ideas. The assumptions underlying each technique should be clearly expressed and carefully described. While the mathematical argument which leads from assumptions to conclusions should be omitted in a book written for the sort of public for which the present book is intended, it is still possible in most cases to explain the fundamental assumptions to the intelligent layman without recourse to any complex mathematical ideas. If space is lacking, it seems desirable to this reviewer to obtain it by reducing or eliminating the considerable detail in which computational techniques are usually described, since misapplication of statistical methods is more likely to be serious than the use of inefficient computing techniques. Much circumlocution could be avoided and considerable clarity and precision

gained if formulae were stated in algebraic symbols, and not described in words and computational procedures designed to be imitated. Omission of mathematics implies to the reviewer's mind the omission of mathematical ideas or arguments of any complexity; the knowledge required to read the elementary algebraic symbolism in which the classical statistical formulae are expressed is now taught at a comparatively early age and is a part of the knowledge of a large section of the population.

The present book contains a number of errors, which are the more serious in that the book is intended for a type of reader who lacks the technical training to be competently critical. A few examples follow:

a) No distinction is made between the population variance and its sample estimate, either in symbolism or in the text (*vide* bottom of page 12).

b) The author fails to realize that a population parameter being estimated is not a chance variable. This occurs in a number of places; for example, on page 14 the author states ". . . the mean has a 95% chance of lying anywhere between 0.18 and −2.38," where the mean referred to is the population mean.

c) The author defines the variance ratio F as the ratio of the larger variance to the smaller. The effect of this is to double the intended size of the critical region under the null hypothesis.

The present book is prefaced by a short introduction written by an official of the British Ministry of Supply, and it is only through his brief acknowledgment that we learn the name of the author. The proper allocation of scientific credit is important not only as a matter of justice, but also because, in the present atomic era, it seems likely that more and more research will be done in large organizations and by scientific teams, and only proper allocation of credit will bring true merit to the fore. A book should bear the name of its author and not that of any administrative bigwig who is his senior in some organizational hierarchy.

On Statistical Treatment of the Results of Parallel Trials With Special Reference to Fishery Research. *H. J. Buchanan-Wollaston* (Principal Naturalist on the staff of the Ministry of Agriculture and Fisheries, working at the Laboratories of the Freshwater Biological Association, Wray Castle, Ambleside, Westmorland, England). Ambleside, Westmorland, England: Freshwater Biological Association of the British Empire, Wray Castle, 1945. Pp. 55. 2s. 8d. Paper. *Two reviews follow:*

REVIEW BY A. M. MOOD
Statistical Laboratory, Iowa State College

FISHERY research enters only incidentally (in some of the illustrative examples) in this booklet which is primarily concerned with tests of significance for 2×2 tables and for samples from binomial populations. The author addresses himself to people with little or no knowledge of statistics, it being his aim to encourage greater use of statistical techniques by workers in fishery research. While such missionary efforts deserve the highest praise,

especially in this case because the author's methodology is excellent, the beginner is warned not to read this booklet for his introduction to statistics. The sentences are complex; the ideas are expressed in very general language; and many technical terms of statistics are used without explanation or definition.

For statisticians, the booklet may be described as an elaboration of Fisher's treatment of 2×2 tables and binomial populations in his *Statistical Methods for Research Workers*. Exact and approximate methods for computing probability levels are fully discussed and illustrated. The author urges research workers to report exact probability levels rather than state merely that P is less than .01 or between .05 and .01 for example. To this end he does not hesitate to recommend quite burdensome arithmetical computations when there is no simpler method of obtaining the exact value. Most statisticians would be less insistent on this point.

A table is provided of the exact probability distributions for 2×2 tables in which one pair of marginal totals are equal. Distributions are given for $n=2, 3, \cdots, 15, 20, 30$, where n is one of the two equal marginal totals. The table would be quite useful to research workers who commonly encounter data of this kind.

The author also presents an interesting technique for combining several experiments in one test of significance when the data are not sufficiently homogeneous that they can be combined directly. The method consists of finding the normal deviates corresponding to the probability levels of the separate experiments, then testing the hypothesis that the deviates are a random sample from a normal population with zero mean and unit variance. This method would be used when the deviates could be given meaningful signs, and would in this circumstance be preferable to Fisher's method of taking minus twice the logarithms of the probability levels to be chi squares with two degrees of freedom.

REVIEW BY MILNER B. SCHAEFER
Instructor in Fisheries, University of Washington

THIS booklet deals with enumeration data, arising from parallel trials, to which may be applied tests of significance based on the fourfold table or on the binomial distribution. Throughout, the result of two parallel trials *is considered as the difference between the number of times a specified* event has occurred in the two trials. Significance is judged by the probability of the occurrence under a null hypothesis of a difference at least as great as that observed. Procedures are given for computing this probability exactly for the fourfold table and for the binomial series. A table of logarithms of the binomial coefficients up to 50 terms is included and greatly facilitates the computations. For the case in which the number of individuals, n, is the same in the two trials, tables of the probability, $P(d)$, for all values of the difference are given for $n \leq 15$, $n=20$, $n=30$. In addition to the exact methods, several methods of approxi-

mating to $P(d)$ are demonstrated for the fourfold table. A method of approximating to the sum of the terms of the binomial series above a given term is presented, based on summing a geometric series, the common ratio of which is estimated from successive terms of the binomial. The testing of the homogeneity of results from several sets of trials, and the combination of results from such trials when they are heterogeneous are briefly treated.

The material is presented almost in the form of a manual of procedures. Each "test" is presented in a separate section, and numerical examples are given in each case. Toward the end of the paper is included a section on "suggestions as to choice of method for testing for significance in particular cases," which will assist the user in selecting tests for his own data. A final section gives examples of problems to which the various methods may be applied. The theoretical basis of the paper is left to eight pages of "theoretical notes" near the end. This method of presentation will facilitate the use of the procedures by biologists and others wishing to employ the tests in the analysis of their own data. However, some unity of subject matter is thereby lost, and it could be wished that the interrelationships of the various sections had been more clearly indicated. For example, the method (of Yates) given in section 6 for computing $P(d)$ in cases where there are over 50 individuals in the trial may be equally well applied to cases with fewer individuals, and, indeed, the formula employed therein is the same as that of sections 1 and 2, but here the factorial products are evaluated by means of the table of binomial coefficients.

The null hypothesis as a basis for tests of significance is discussed in the introduction, and is again emphasized in connection with the various procedures. This, together with a separate section "on the philosophical basis of tests for significance," should give readers with a limited background of statistical theory a valuable insight into the logic underlying the methods of this paper.

The examples employed in illustrating the procedures are drawn from various problems arising in fisheries studies, but the methods will be found generally applicable in other fields.

Statistical Methods in Quality Control. *P. C. Clarke* and *G. R. Armstrong.* (Engineering Department, Hunter Pressed Steel Co.) Engineering Data Book Section 20. Lansdale, Pa.: Hunter Pressed Steel Co., 1945. Pp. 100 (pagination irregular).

Review by G. R. Gause
Bell Telephone Laboratories, Inc., New York, N. Y.

This book presents in looseleaf style a description of the quality control methods which have been found useful by the Hunter Pressed Steel Company. The pages are exact reproductions of data sheets which were prepared by the company for their inspectors and inspector trainees, being in large part specific operational procedures to be followed by the inspection per-

sonnel. As such, the book lacks some of the continuity which is desirable in a textbook. However, the simplicity of presentation and the wealth of practical illustrations should make the book of much value both to the young engineer who is starting out in quality control work and to the teacher of industrial statistics. To the young engineer, this book, supplemented by one of the standard texts, should be invaluable in bridging over the gap between the more abstract considerations and the practical applications. To the teacher, it provides both techniques for the presentation of some of the fundamental concepts of quality control and also many numerical examples of practical applications.

The first part deals with frequency distributions. The concept of a frequency distribution is very cleverly and lucidly developed through a series of simple examples and illustrations. Detailed procedures are given for calculating $\overline{X}$ and σ and for estimating fraction defective from computed values of $\overline{X}$ and σ. The method given for calculating $\overline{X}$ and σ from the observed fractions defective in two tails of a distribution is, as far as this reviewer's knowledge goes, here published the first time. The method should be of practical value in many instances. However, as the author undoubtedly realizes, the method given can be dangerous if not handled with caution. Had the book been prepared for the general public rather than for use by a particular company under careful supervision, it would have been desirable to include some cautioning remarks as to the dangers inherent in making predictions from such calculations, particularly when the observed fractions defective are both in the neighborhood of .50. The illustrations of the effect upon frequency distributions of such things as tool wear, heat treating operations, and other factors should be particularly helpful to the practical man who is trying to understand what frequency distributions have to do with making a satisfactory product.

The second portion of the book deals with correlation between two quality characteristics. The treatment omits completely all the mathematics of correlation and treats the subject in a simple graphical manner such as is so helpful to the shop man in quickly detecting pronounced relationships and in presenting these relationships in an obvious manner. Illustrations are given as to practical applications which resulted in the solution of quality problems in the shop.

The last portion of the book deals with control charts. In general, the methods presented are those of the American Standards Association with several additions which have been found useful by the authors. One of these additions is the use of "reject limits." The "reject limits" are placed inside the specification limits by a factor depending upon sample size and observed average range. Sample averages falling outside these "reject limits" are considered by the company as sufficient cause for stopping production and making adjustment. The limits are of the same nature as those described by E. H. Sealy in his book *A First Guide to Quality Control for Engineers*, in which book the limits are termed "modified control limits." It is interesting

to note that the Hunter Pressed Steel Company has found such limits to be of considerable value in their day-to-day work. Another addition is the "control gaging" technique. This is provided as a substitute for the routine keeping of control charts by variables and consists in setting go not-go gages at the control limits for averages of samples of five. The inspection procedure then consists in gaging with the "control gage" a sample of five pieces at periodic intervals. If more than two pieces fail either the go or the not-go gage, an adjustment of machinery is required. It is pointed out that this procedure eliminates the measuring of individual pieces, the calculation of averages and ranges, and the plotting of points, yet furnishes a practical means of control where more refined methods are not justified. It would be interesting to know if the company has tried setting the "control gages" inward from the specification limits by a factor of observed range rather than outward from the center of the specification tolerance range. Such a setting of the control gages would correspond to the "reject limits" which apparently have been found to be so successful and would automatically take care of situations such as that described on page 114 in which there is a relatively small range but a pronounced trend in averages. Another addition to techniques described in the ASA Standards is a procedure to be followed on short-run jobs which cannot be treated as a continuous production situation. The procedure given illustrates the fact that applications of quality control methods need not be confined to continuous mass production.

A number of reproductions of control charts prepared on the job are given at the end of the book.

The Recovery of Inter-Block Information in Quasi-Factorial Designs with Incomplete Data. Commonwealth of Australia, Council for Scientific and Industrial Research. *E. A. Cornish* (Officer-in-Charge, Section of Mathematical Statistics, the Council). East Melbourne, C. 2, Australia: the Council, 314 Albert St. Gratis. Paper.

1, Square, Triple, and Cubic Lattices. Bulletin No. 158. 1943. Pp. 22.
2, Lattice Squares. Bulletin 175. 1944. Pp. 19.

Review by C. H. Goulden
Officer-in-Charge, Dominion Laboratory of Cereal Breeding
Winnipeg, Manitoba, Canada

These papers are a continuation of a series in which the author presents correct methods of analysis for incomplete block or quasi-factorial experiments when the data are incomplete. In order to familiarize the reader with the terms and symbols used, the author presents first the fundamentals of the theory underlying each type of experiment with which he deals. The formulas for the estimation of missing values are then presented clearly and detailed instructions are given for the solution of the rather more perplexing problem of making adjustments to sums of squares for use in tests of significance. The methods described, together with those given in previous papers, bring up to date the methods of analysis for quasi-factorial designs for the three situations that arise in the use of experiments of this type,

(*a*) when there are no real differences between the blocks and consequently no adjustments are required, (*b*) when the differences between the blocks are very large and full adjustments are made, and (*c*) when the differences between the blocks are intermediate between the extremes as in (*a*) and (*b*) and the magnitude of the adjustments made are based on an estimate of the block differences determined from the data of the experiment. Actual examples are given.

Lectures on Probability and Statistics. *Edward L. Dodd* (Late Professor of *Actuarial Mathematics, University of Texas*). Austin, Texas: University of Texas Press, 1945. Pp. 44. Gratis. Paper.

REVIEW BY THOMAS N. E. GREVILLE
Actuarial Mathematician, Bureau of Census, Washington

THIS booklet contains three lectures given by the late Professor Dodd on the theory of means or averages, together with a biographical sketch of the author. The lectures do not pretend to cover the subject exhaustively, but are of a semi-popular nature, seeking to indicate the general nature of the subject and to stimulate interest. For those who desire to read further, references are given to all the important papers. Much of the discussion in the second and third lectures centers around certain highly generalized defitions of a mean or average, of which the familiar types, such as arithmetic, geometric, and harmonic means, median, mode, and midrange, are special cases. To the uninitiated this is likely to appear a highly academic study, but the author shows by several well chosen examples taken from the fields of statistics, compound interest, and actuarial science that a number of unusual or peculiar types of means or averages are in common use for certain purposes. In the third lecture, mention is made of a number of specific properties which have been used to classify the various types of means. For example, a mean of a set of numbers is said to be internal if it lies between the greatest and the smallest of the numbers, or at worst equals the maximum or minimum value; it is said to be associative if it remains unchanged when all the numbers in any subset (taken at pleasure) are replaced by the mean for that subset. In the opinion of the reviewer, these lectures admirably fulfil their purpose of popularizing and arousing interest in the subject.

An Introduction to Industrial Statistics and Quality Control. *Paul Peach* (Industrial Statistician, Institute of Statistics, University of North Carolina, Raleigh). Raleigh, N. C.: L. B. Phillips Company, 1945. Pp. ix, 150. Out of print. Paper. (*Revised edition in press, Edwards & Broughton, Raleigh, N. C.*)

REVIEW BY J. H. CURTISS
*Lt. Comdr., USNR, Bureau of Ships, Washington, D. C.**

THIS book is a preliminary edition, in mimeographed form, of a textbook on statistical quality control which Mr. Peach is now preparing. While the

* The opinions expressed in this article are those of the author and are not to be construed as reflecting the views of the Navy Department or the Naval Service at large.

number of copies obtainable is limited, the scarcity of satisfactory working texts in this field seems to justify paying especial attention to the present work.

The immediate background of the book was Mr. Peach's experience as a teacher in the eight-day quality control courses given during the war in various industrial centers by a unit of the WPB under Mr. Holbrook Working. The effect of these courses in spreading the word on statistical quality control has been extremely significant. Although this book goes considerably beyond the subject matter of the eight-day courses, and changes the order of presentation of topics usually employed in them, it is nevertheless a fairly obvious outgrowth of the schedule employed for the courses. It thus derives a certain importance from being the first formal write-up in book form of the subject matter of the WPB program.

This is not meant to impute any lack of originality to the author. As stated above, many topics are included which were not treated in the WPB courses, and some of the devices employed seem to have originated with the author.

The order of presentation departs somewhat from that of the WPB courses in that acceptance sampling is presented before control charts. From the theoretical viewpoint, this is certainly the way to go about the matter, but some teachers of quality control prefer to put the control charts first. The presentation of the rationale of acceptance sampling is very well done. A system for selecting single sampling plans by attributes on the basis of the ratio p_2/p_1 is given, which is essentially the same as that employed in the earlier paper by Dodge and Romig on acceptance sampling.[*] The reviewer would have supposed that it would have been advisable to defer the details of selecting sampling plans in this way to a later portion of the book, but a personal contact with a correspondence-school student who is using the text has revealed that he was not particularly discouraged by encountering such complications at an early stage in the work.

Acceptance sampling by attributes leads naturally to the p-chart, although the distinction is not carefully made between the control of quality of the process and of the individual lot. The $\bar{x}$ and R chart appears next, after some discussion of the standard deviation. A novelty then appears in the form of a quality control method for short runs. It essentially involves the use of the sample mean and standard deviation to control the proportion defective, assuming a normal distribution. It might be advisable to use the non-central t distribution for this purpose; perhaps this modification will appear in the final edition.

At this point in the book, a certain evidence of haste seems to make its appearance in the organization of material. The list of contents from here on in reads as follows: The Chart for c; Multiple Sampling; AOQL Inspection; Acceptance Inspection by Variables; Specifications and Tolerances; Trend Control; Gages, Sampling, and Scientific Measurements; An Introduction

[*] H. F. Dodge and H. G. Romig, "A Method of Sampling Inspection," *Bell System Technical Journal*, 8:613–31 O '29.

to the Analysis of Variance; General Remarks on Raw Materials Acceptance; Organizing the Quality Control Department. Each topic receives a good, although brief treatment, but the net effect is a little bewildering. Undoubtedly the problem of organization and emphasis will be solved more satisfactorily in the final edition.

There are very few really misleading statements. Perhaps the reviewer's chief criticism would be that no connection is established between the *a priori* definition of probability used to introduce the topic and the frequency ratio definition which underlies much of the applications. A few statements need a little qualification: on page 9 it is stated without mention of independence that the variance of a sum is equal to the sum of the variances; on page 18 it is alleged that it is impossible to gage the quality of a lot by a sample (which denies the theory of statistical estimation); and on page 75 it is observed that the theory of double sampling is incomplete and existing tables are empirical, which is perhaps a little misleading.

These minor blemishes do not really mar a very accurate and readable book. Indeed, this work is exceptionally well written from start to finish. It is one of the best elementary expositions of statistical methods which the reviewer has ever encountered. The author is apparently a very fine teacher in his chosen field. He shows a grasp of the theoretical background which is particularly commendable in view of the fact that much of the attention in the exposition is necessarily centered on keeping the text strictly down to earth.

The Analysis of Cubic Lattice Designs in Varietal Trials. *I. F. Phipps, A. T. Pugsley, S. R. Hockley,* (Waite Agricultural Research Institute, University of Adelaide, Adelaide, South Australia); *E. A. Cornish* (Officer-in-Charge, Section of Mathematical Statistics, Council for Scientific and Industrial Research). Commonwealth of Australia, Council for Scientific and Industrial Research, Bulletin No. 170. East Melbourne, Australia: Council for Scientific and Industrial Research, 314 Albert St., 1944. Pp. 41. Gratis. Paper.

Reviewed by Gertrude M. Cox
Director, Institute of Statistics
University of North Carolina, Raleigh

The authors present the construction of the unbalanced cubic lattice designs and discuss their value in favorable terms. Their own summary is concise and to the point: "The paper describes in detail the procedure to be adopted for recovering inter-block information in cubic lattice designs, and the computations which are required for this full analysis are illustrated by a numerical example."

This bulletin follows closely Yate's original description illustrating the recovery of inter-block information in cubic lattice designs published in 1939. The authors are correct in their statement that reprints of Yate's article are not readily available. A volume of the *Annals of Eugenics* cannot be used extensively by computers in a laboratory. Therefore, a bulletin such as this one provides a useful guide.

The formulae in both articles are the same but the notation differs in minor details. To solve block components (a), (b), and (c) a series of separate tables 2, 3, 4 and 5 are constructed. This makes the computing instructions much easier to follow than those given by Yates.

After giving the logic of the computing involved to secure component (c), the contribution from tables 3a, 3b and 3c could all be thrown into one solution, thus making the process straight forward for computers.

The conditions specified for using these block components for the estimation of w′ should be noted.

(1) The mean squares corresponding to these three components of the block sum of squares have different expectations which are independent of the number of times the cubic lattice is replicated.

(2) If the whole cubic lattice is replicated twice, to estimate w' use component (a) and the additional $3(p^2-1)$ degrees of freedom for differences between blocks containing the same varieties. The resultant mean square is based on $3(p-1)(p+2)$ degrees of freedom with expectation $pA+B$.

(3) When inter-block mean square is less or equal to the intra-block mean square analyze as a randomized block design.

The formula for the variance of differences between adjusted varietal means can be simplified considerably. When making the test of significance and also when solving for values to use for adjusting varietal yields, λ and μ were used. That is,

$$\lambda = \frac{w - w'}{w + 2w'} = 0.4020, \qquad \mu = \frac{w - w'}{2w + w'} = 0.2807$$

let

$$\lambda' = \frac{\lambda}{p^2} = 0.01608, \qquad \mu' = \frac{\mu}{p^2} = 0.01147$$

then

$$V_m = \frac{2Ei}{3r}\left[1 + \frac{p^2}{p^2+p+1}\left\{6\lambda' + 3(p-1)\mu'\right\}\right]$$

$$V_m = \frac{2(.3682)}{3}\left[1 + \frac{25}{31}\left\{6(.01608) + 3(4)(.01147)\right\}\right]$$

$$V_m = \frac{2(.3682)}{3}(1.1888) = .2918$$

which is easier to solve than the formula as given on page 37. Also, note that $2Ei/3r$ is the regular formula for the variance of the differences between varietal means. The 1.1888 is the factor introduced by confounding.

Then, the effective error mean square per unit is $(.3682)(1.1888) = .4377$. This value compared with the error mean square (residual) secured from the ordinary randomized block analysis gives the efficiency of the cubic lattice design.

$$100 \times \frac{.5120}{.4377} = 117\%$$

Again, this is a much shorter process than that given on pages 38 and 39. In fact, table 9 gives the efficiencies accurately enough for most purposes. In this example

$$\frac{w}{w'} = \frac{2.7150}{.9002} = 3.017.$$

Reading from the table gives 116.8%.

From the standpoint of the applied statistician, the presentation in this bulletin is excellent. At each step, the mathematical discussion and formulae are followed by the corresponding numerical results from the example.

Sequential Analysis of Statistical Data: Applications. Prepared by the *Statistical Research Group, Columbia University* for the Applied Mathematics Panel, National Defense Research Committee, Office of Scientific Research and Development. SRG Report 255, Revised; AMP Report 30.2R, Revised. New York: Columbia University Press, September 1945. Pp. vii, 17; iv, 80; v, 57; iii, 25; iii, 18; iii, 30; ii, 41. $6.25. *Two reviews follow:*

Review by Henry Scheffé

Associate Professor of Engineering

University of California at Los Angeles

This work constitutes a manual on the application of the Wald sequential probability ratio test to certain specific problems. In this test the observations are taken one at a time, and after each observation one of the following three actions is adopted: (1) the lot or process being inspected, or the statistical hypothesis being tested, is accepted, (2) it is rejected, (3) judgement is withheld and it is decided to take another observation. In cases (1) and (2) no further observations are taken. The present exposition is mostly in the language of inspection, although connections with experimentation are also noted. The descriptions of the properties of sequential tests are clear, and the instructions for their use are detailed; little background in the way of statistics or mathematics is demanded of the reader. The clarity is improved by the omission of some of the qualifications that a mathematician might be impelled to add; for example, the advantages of sequential testing are well summarized for non-mathematical readers in the following statement: "Thus, conspicuously good lots are quickly accepted, conspicuously bad lots are quickly rejected, and extensive inspection is needed chiefly by lots of doubtful quality—which is as it should be."

In every case treated here the probability of acceptance, when a given sampling plan is used, depends on a single unknown parameter θ. The graph of this probability is called the *OC* (operating characteristic) curve. The average number of observations required to reach action (1) or (2) above also depends on θ; its graph is called the *ASN* (average sample number) curve. The customer for the test chooses two distinguished values of θ, say θ_1 and θ_2, and ordinates of the *OC* curve at these values; ordinarily one of these ordinates will be chosen near unity and the other near zero. Considera-

tions governing the choice of θ_1, θ_2, and the corresponding ordinates are discussed in the introductory section 1. The *OC* curve which passes through the two chosen points is everywhere rising or everywhere falling, except in section 5 (the case of "two-sided alternatives"), where it is symmetrical in the line $\theta=\theta_1$ and falling for all θ greater than θ_1. The problems treated in sections 2 to 6 are the following:

2. θ is the unknown parameter (population proportion) of a binomial population.

3. There are two binomial populations with unknown parameters p_1 and p_2. In the first treatment θ is the "odds ratio," that is, the quotient of $p_1/(1-p_1)$ by $p_2/(1-p_2)$. In the second treatment θ is the "angular difference,"* that is, $\theta=\arcsin\sqrt{p_1}-\arcsin\sqrt{p_2}$.

4. θ is the mean of a normal population of known variance.

5. Ditto.

6. θ is the standard deviation of a normal population. In the first treatment the mean of the population is assumed known, in the second treatment, unknown (but the same for all θ).

Each of these sections describes both a graphic and a tabular procedure for analyzing the data, and each contains at least one numerical illustration. Except for section 5, each section contains a discussion of the *OC* and *ASN* curve; in this and other ways there is a considerable enlargement of the original "restricted" edition. Practically all the errors noted in that edition have now been eliminated. An extra section consisting of appendices treats among other matters the device of truncating the previously described sequential tests. This device, which is of great practical importance, consists of choosing beforehand an upper bound for the sample size, and whenever this size is reached without actions (1) or (2) having been taken in the usual way, one or the other is then taken, depending on the observations accumulated.

The organization and format are unusual. While the reviewer likes to see experiments in organizing and binding statistical material, he regrets finding this one somewhat unfortunate. The seven sections are separately bound in pamphlet form, stapled, not stitched, and they come in a ring binder, the whole arrangement being rather flimsy. When the pamphlets are used in the binder the book is hard to close unless a suitable process of shaking down is first used. The text is reproduced from typewritten copy by a photographic process. An effort has been made to make each of sections 2 to 6 practically self-contained, resulting in a large amount of repetition.

REVIEW BY B. L. WELCH
Department of Statistics, University College, London

THIS book is a revision and extension of work which was issued earlier as an O.S.R.D. report with restricted circulation. Its publication and

* The theoretical justifications of some innovations in connection with the angular transformation are not given. They will appear in *Techniques of Statistical Analysis for Selected Problems in Science, Engineering, Production, and Management* by the Statistical Research Group, Columbia University, to be published by McGraw-Hill Book Company in 1946.

general circulation to a wider public at this moment is welcome. In conjunction with the more theoretical papers on the same subject, which have now appeared in various statistical journals, it makes available a very clear account of a type of statistical procedure which has been strongly advocated in several quarters. It should be noted, however, that although sub-titled "Applications," the present book is only rightly so-called in the sense that it makes clear how to operate certain sampling procedures and how to analyse the observations resulting therefrom, assuming all along that these procedures are the ones most appropriate to the questions at issue. Except, perhaps, by implication, it scarcely touches the problem of indicating in practice the types of job for which the methods suggested are genuinely preferable to others which might be used.

The term "Sequential Analysis" has on occasion been applied to cover the analysis of data arising in all situations where sample sizes are not fixed in advance, but where the decision to stop sampling depends on the outcome of observations as they are made. It is not with such a wide meaning, however, that the term is used in the present book. The procedures described are the particular ones developed by Wald and his co-workers at Columbia University and any remarks in this review must therefore also be taken to refer only to these.

To illustrate the type of approach adopted we may instance one of the sections of the book, which deals with the case where, in the material sampled, the character in question is normally distributed with known standard deviation but unknown mean, and where a sampling plan and a rule for accepting or rejecting batches is required with the following properties: The chance of *accepting* material for which the true mean is m_1 should not exceed some specified small risk α; on the other hand the chance of *rejecting* material for which the true mean is m_2 should not exceed some specified small risk β. The suggested sampling procedure is then this: As every extra individual is added to the sample the decision is made whether to accept the batch sampled, reject it or take another observation. The decision is made on the basis of the running total, Σx, of the observations accruing to date, and sampling is continued till a decision one way or the other is obtained.

The claim that is made for this procedure as against a procedure that uses a sample number fixed in advance, is that it can satisfy the conditions laid down (i.e., give practically the same operating characteristic) with an appreciably smaller *average* sample size. More precisely it can be shown that, if the true mean m is in the vicinity either of m_1 or m_2, then the average amount of sampling will be appreciably less than if a fixed sample scheme were being used. However, if m is about half-way between m_1 and m_2, no such reduction in sample size will be obtained.

Results of this nature are interesting, but inconclusive if no attempt is made to relate them to the values of m likely to occur in a given practical case, and it is at this point that the present book proves somewhat disappointing. A basic criticism which has been made of Wald's work is that fre-

quently m will lie about half-way between m_1 and m_2 and that therefore appreciable average reductions in sampling size will not happen. Further, that even when m is either in the vicinity of m_1 or m_2 (so that the material sampled is in some sense abnormal) we may not be concerned so much with accepting or rejecting it, as with estimating its properties. For such purposes the smaller sample sizes then arising from the sequential method may be a positive disadvantage.

Whatever substance there may be in such criticisms must depend of course on circumstances and further discussion in abstract terms is probably not very profitable. In a book dealing with the "applications" of a statistical method, we might, however, have expected that more space would have been devoted to anticipating possible objections of this nature even if, in the opinion of the authors, such objections are not of great importance.

Other sections of the book deal in a very similar manner, with problems of controlling other features of the batch, e.g., proportions defective, variabilities, etc. Each section is bound separately in pamphlet form, and the whole series assembled in a ring type folder. This format is adopted so that the relevant section for any particular job can be abstracted and so be more conveniently available to anyone using it often. For library purposes a further edition of the book bound up in the usual manner would possibly be more convenient.

PUBLICATIONS RECEIVED

Statistical methodology publications received are listed elsewhere in this issue under the section "Statistical Methodology Index."

Business Literature, 1943–1944: With Cumulated Index. Newark, N. J.: Newark Public Library, 1944. Pp. 236.

Capital Goods Industries and Postwar Taxation. Chicago, Ill.: Machinery and Allied Products Institute (221 N. LaSalle St.), 1945. Pp. 22. Paper.

Commercial Policy in the Post-War World: Report of the Economic and Financial Committees. Series of League of Nations Publications, II, Economic and Financial, 1945.II.A.7. New York: Columbia University Press, 1945. Pp. 124. $1.25. Paper.

The Financial Record of the Electric Utility Industry, 1937–1944: A Summary Compiled from Annual Reports of the Federal Power Commission Entitled "Statistics of Electric Utilities in the United States." Washington: Federal Power Commission, November 15, 1945. Pp. ii, 14. Gratis. Paper.

Ice Cream Sales Index, January 1st Through December 31st, 1944: An Analysis of Ice Cream Sales in 1944 Compared with 1943 for the United States and Canada. Special Bulletin No. 60. Washington: International Association of Ice Cream Mfgs. (1105 Barr Bldg.), July 1945. Pp. 54. $1.00. Paper.

Population Trends in Southwestern Pennsylvania, 1910–1944. Prepared for the Allegheny Conference on Community Development. Pittsburgh 22, Pa.: Bureau of Social Research, Federation of Social Agencies of Pittsburgh and Allegheny County (519 Smithfield St.), October 1945. Pp. vii, 33. $0.50. Paper, mimeographed.

Public Investment and Capital Formation: A Study of Public and Private Investment and Outlay, Canada 1926–1941. Ottawa, Canada: Dominion-Provincial Conference on Reconstruction, 1945. Pp. 128. Paper.

Social Facts About Pittsburgh and Allegheny County: Vol. II, Pittsburgh Study Areas. Pittsburgh 22, Pa.: Bureau of Social Research, Federation of Social Agencies of Pittsburgh and Allegheny County, 1945. Pp. 64.

Union Labor in California, 1944. Report No. 1. San Francisco, Calif.: Division of Labor Statistics and Research, Department of Industrial Relations, State of California, October 1945. Pp. 46. Paper.

Year Book of Labour Statistics, 1943–44, Eighth Issue. Washington 6: Washington Branch, International Labour Branch (734 Jackson Place), 1945. Pp. xx, 205. Cloth, $3.00; paper, $2.00.

The Economics of the Pacific Coast Petroleum Industry: Part 2, Price Behavior and Competition. *Joe S. Bain.* Publications of the Bureau of Business and Economic Research, University of California. Berkeley, Calif.: University of California Press, 1945. Pp. xv, 438. $6.00.

Group Health Insurance and Sickness Benefit Plans in Collective Bargaining. *Helen Baker* and *Dorothy Dahl.* Research Report Series, No. 72. Princeton, N. J.: Industrial Relations Section, Princeton University, 1945. Pp. 94. $1.50. Paper.

The Development of Mathematics, Second Edition. *E. T. Bell.* New York: McGraw-Hill Book Co., Inc., 1945. Pp. xiii, 637. $5.00.

The British Balance-of-Payments Problem. *Arthur I. Bloomfield.* Essays in International Finance, No. 6 Princeton, N. J.: International Finance Section, Princeton University, 1945. Pp. 28. Paper.

Bioenergetics and Growth: With Special Reference to the Efficiency Complex in Domestic Animals. *Samuel Brody.* A Publication of the Herman Frasch Foundation. New York: Reinhold Publishing Corp., 1945. Pp. xii, 1023. $8.50.

Agricultural Adjustment and Income. *Arthur C. Bunce, William H. Fisher,* and *Earle L. Rauber.* Postwar Economic Studies, No. 2. Washington: Division of Administrative Services, Board of Governors of the Federal Reserve System, October 1945. Pp. iv, 60. $0.25. Paper.

Expenses and Profits of Limited Price Variety Stores in 1944: Chains and Independents. *Elizabeth A. Burnham.* Bureau of Business Research, Bulletin No. 123. Boston, Mass.: Division of Research, Harvard Business School, 1945. Pp. vi, 35. $1.50. Paper.

Effect of Federal Taxes on Growing Enterprises. *J. Keith Butters* and *John*

Lintner. Boston 63, Mass.: Division of Research, Harvard Business School, 1945. Pp. ix, 220. $3.00.

Freedom From Want After the War: A Postwar Program for the Reduction of Idleness and Want. *Wayne F. Caskey*. LaSalle, Ill.: the Author (1500 St. Vincents Ave.), 1945. Pp. 66. $0.75. Paper.

Working Conditions and Employee Services. *B. J. Cohen* and *M. M. Towy-Evans*. London, W.C.2: Institute of Labour Management (Aldwych House), 1945. Pp. 87. 2s. Paper.

A Tax Program for a Solvent America. *The Committee on Postwar Tax Policy*. [A condensation of the book of the same title published by the Ronald Press.] New York: the Committee (50 West 50th St.), 1945. Pp. 48. Gratis. Paper.

Apprentice Practice in the United States: As Revealed by an Analysis of Workable Apprenticeship Programs in American Industry. *Eugene Danaher*. Business Research Series, No. 3. Stanford University, Calif.: Graduate School of Business, Stanford University, 1945. Pp. iv, 60. $0.75. Paper.

Wages of Agricultural Labor in the United States. *Louis J. Ducoff*. United States Department of Agriculture, Technical Bulletin No. 895. Washington: U. S. Government Printing Office, July 1945. Pp. 127. $0.25. Paper.

Argentina's Agricultural Exports During World War II. *Pavel P. Egoroff*. War-Peace Pamphlets, No. 8. Stanford University, Calif.: Food Research Institute, Stanford University, November 1945. Pp. iv, 52. $0.50. Paper.

Bilateralism and the Future of International Trade. *Howard S. Ellis*. Essays in International Finance, No. 5. Princeton, N. J.: International Finance Section, Princeton University, 1945. Pp. 22. Paper.

Tax Relief Under Section 722. *Paul W. Ellis*. Conference Board Reports, Studies in Business Policy, No. 9. New York 17: National Industrial Conference Board, Inc. (247 Park Ave.), 1945. Pp. 48. Paper.

Labor Savings in American Industry, 1899–1939. *Solomon Fabricant*. Occasional Paper 23. New York 23: National Bureau of Economic Research (1819 Broadway), November 1945. Pp. iv, 51. $0.50. Paper.

100% Money: Designed to Keep Checking Banks 100% Liquid; to Prevent Inflation and Deflation; Largely to Cure or Prevent Depressions; and to Wipe Out Much of the National Debt, Third Edition. *Irving Fisher*. New Haven, Conn.: City Printing Co., 1945. Pp. xxii, 257. $1.00.

Planning for Jobs: Proposals Submitted in the Pabst Postwar Employment Awards. Edited by *Lyle Fitch* and *Horace Taylor*. Philadelphia, Pa.: Blakiston Co., 1946. Pp. xxii, 463. $3.75.

Poona: A Socio-Economic Survey, Part I, Economic. *D. R. Gadgil*. Publication No. 12. Poona 4, India: Gokhale Institute of Politics and Economics, 1945. Pp. x, 300. Rs 15 or 30s.

Jobs, Production, and Living Standards. *E. A. Goldenweiser*, *Everett E. Hagen*, and *Frank R. Garfield*. Postwar Economic Studies, No. 1. Washington: Division of Administrative Services, Board of Governors of the Federal Reserve System, August 1945. Pp. iv, 85. $0.25. Paper.

Cleveland Market Data Handbook: 1945 Edition. *Howard Whipple Green*. Cleveland 15, Ohio: Real Property Inventory of Metropolitan Cleveland (1001 Huron Road) 1945. Pp. 69. Paper.

From Bretton Woods to Full Employment. *Dag Hammarskjöld*. Supplement A to Svenska Handelsbanken's Index, December 1945. Pp. 21. Paper.

The Social Framework of the American Economy: An Introduction to Economics. *J. R. Hicks* and *Albert Gailord Hart*. New York: Oxford University Press, 1945. Pp. xvii, 201. $2.50.

Financing American Prosperity: A Symposium of Economists. Edited by *Paul T. Homan* and *Fritz Machlup*. New York 18: Twentieth Century Fund (330 West 42 St.), 1945. Pp. xi, 508. $3.00.

International Trade and Domestic Employment. *Calvin B. Hoover*. Committee for Economic Development Research Study. New York: McGraw-Hill Book Co., Inc., 1945. Pp. xii, 177. $1.75.

Marketing and Manufacturing Margins for Textiles. *L. D. Howell*. United States Department of Agriculture, Technical Bulletin No. 891. Washington: U. S. Government Printing Office, March 1945. Pp. 148. $0.20. Paper.

Method of Indexing Provisions of Collective Agreements, Second Edition. *May E. Jamison*. Bulletin No. 3. Pasadena 4, Calif.: Industrial Relations Section, California Institute of Technology, 1945. Pp. iv, 16. $1.00. Paper.

World Rubber and Its Regulation. *K. E. Knorr*. A Publication of the Food Research Institute, Stanford University, Calif.: Stanford University Press, 1945. Pp. x, 265. $3.00.

A Short History of Labour Conditions Under Industrial Capitalism: Vol. 3, Part I, Germany, 1800 to the Present Day. *Jürgen Kuczynski*. London, W.C.1: Frederick

Muller Ltd. (20 Great James St.), 1945. Pp. 208 9s. 0d.

Labor Fact Book 7. *Labor Research Association.* New York 10: International Publishers (381 Fourth Ave.), 1945 Pp. 208. $1.00.

World Politics Faces Economics: With Special Reference to the Future Relations of the United States and Russia. *Harold D. Lasswell.* Committee for Economic Development Research Study. New York: McGraw-Hill Book Co., Inc., 1945. Pp. x, 108. $1.25.

Debtor and Creditor Countries: 1938, 1944. *Cleona Lewis,* assisted by *Louise Bebb.* Washington: Brookings Institution, 1945. Pp. vii, 99. $0.75. Paper.

The Forest and Forest Industries of Norrland. *Gunnar Löwegren.* Supplement to Svenska Handelsbanken's Index, September 1945. Pp. 63 Paper.

Prospects and Problems in Aviation: A Series of Papers Presented at the Chicago Forum on Aviation, Organized and directed by *Leverett S. Lyon* and *Lewis C. Sorrell.* Chicago 2, Ill.: Chicago Association of Commerce (1 N. LaSalle St.), 1945. Pp. v, 212. $2.00. Paper.

Operating Results of Department and Specialty Stores in 1944. *Malcolm P. McNair.* Bureau of Business Research, Bulletin No. 122. Boston, Mass.: Division of Research, Harvard Business School, 1945 Pp. vi, 42. $2.50. Paper.

Southern State and Local Finance Trends and the War. *James W. Martin.* Papers of the Institute of Research and Training in the Social Sciences, Vanderbilt University, No. 8; Bulletin of the Bureau of Business Research, University of Kentucky, No. 10. Nashville, Tenn.: Vanderbilt University Press; Lexington, Ky.: University of Kentucky Press, March 1945. Pp 106. $0.50. Paper.

Labor Policy of the Federal Government. *Harold W. Metz.* Washington: Brookings Institution, 1945. Pp ix, 288. Paper.

Should Price Control Be Retained? *Harold G. Moulton* and *Karl T. Schlotterbeck.* Washington: Brookings Institution, 1945. Pp. iv, 43. $0.50. Paper.

The Growth of Personnel Management in Great Britain During the War, 1939–1944. *G. R. Moxon.* Supplementary Pamphlet No. 2. London, W.C.2: Institute of Labour Management (Aldwych House), 1945. Pp. 32. 1s. Paper.

The World's Hunger. *Frank A. Pearson* and *Floyd A. Harper.* Ithaca, N. Y.: Cornell University Press, 1945. Pp. vii, 90. $1.50.

Rice in the Western Hemisphere: Wartime Developments and Postwar Problems. *V. D. Wickizer.* War-Peace Pamphlets, No. 7. Stanford University, Calif.: Food Research Institute, Stanford University, June 1945. Pp. vi, 48. $0.50. Paper.

STATISTICAL METHODOLOGY INDEX, NO. 3

A QUARTERLY GUIDE TO CURRENT LITERATURE

Edited by
OSCAR KRISEN BUROS
Rutgers University

This bibliographic service presents statistical methodology literature (articles, books, theses, and chapters) published in 1945 to date. Anonymous references are listed first; other references follow in alphabetical sequence by first-named authors. Volume number, issue number (in parentheses), pagination, and date of issue are given for each journal reference. Stars indicate books, theses, and pamphlets; asterisks indicate publications seen by the editor of this INDEX. Authors of papers on statistical methodology are invited to send reprints to Oscar K. Buros, Rutgers University, New Brunswick, New Jersey in order to facilitate prompt and accurate listings. Foreign-language papers should be accompanied by English translations of titles. Information is desired about references omitted.

Appliance-Saturation Factors: E.R.A. Communication on Measurement by Means of Sampling. *Electrician* 135(3514): 345-6 O 5 '45.* [269

Correlation Methods Applied to Steel Problems. *Blast F & Steel Pl* 33(11):1409-10 N '45.* [270

A Double-Sampling Inspection Scheme. *Prod & Eng B* 4(23):350-5 S-O '45.* [271

★Federal Dimensional Quality Control Primer, Revised Edition. Providence 1, R. I.: Federal Products Corp. (1144 Eddy St.), 1946. Pp. 38. Paper. One to six copies, gratis; seven or more copies, $0.25 per copy.* *To be reviewed.* [272

Indian Statistical Institute Thirteenth Annual Report: 1944-1945. *Sankhya* 7(2): 217-36 N '45.* [273

Indian Statistical Institute Twelfth Annual Report: 1943-44. *Sankhya* 7(1): 107-20 Ag '45.* [274

Quality Control. *Monetary Times* 113(5): 29-31+ My '45.* [275

Reviews of Books on Statistical Analysis, Quality Control, etc. *ASTM B* (138):60-2 Ja '46.* [276

Saturation Factors of Domestic Electric Appliances. *Engineering* 160(4161):202 O 12 '45.* [277

★Symposium of Papers on Statistical Quality Control. Birmingham 3, England: Quality Control Panel, Birmingham District Production Committee, Ministry of Production, C. M. L. Building, Great Charles St., [March 1945]. Pp. 70. Gratis. Paper.* For individual papers, see 278, 292, 326, 401, 437, 440, 454, 464. *Reviewed in this issue.* [278

Training in Statistics at the University of Minnesota. *Biometrics B* 1(5):65 6 O '45.* [279

Wartime Analysis Method [Sequential Analysis] May Aid Insurance Companies: Cuts Data Needed for Spot-Checking by About 50%. *Nat Underw* 50(2):9 Ja 10 '46.* [280

On the Arithmetic and the Geometric Means From a Type III Population. S. Janardana Aiyer (Department of Statistics, University of Travancore, India). *Math Student* 13(1):10 5 Mr '45.* [281

Quality Control Through Product Testing. P. L. Alger (Staff Assistant to the Vice-President in Charge of Design Engineering, General Electric Co., Schenectady, N. Y.). *Elec Eng* 65(1):11-2 Ja '46.* [282

Lognormal Distributions. Letters. Percival Allen (Department of Geology, University of Reading); S. C. Pearce (East Malling Research Station, East Malling, Maidstone, Kent, England); and J. H. Gaddum (University of Edinburgh). *Nature* 156(3973): 746-7 D 22 '45.* [283

Statistical Methods in Quality Control: 6-8. American Institute of Electrical Engineers, Subcommittee on Educational Activities (J. Manuele, H. F. Dodge, A. I. Peterson, and R. E. Wareham). *Elec Eng* 64(11, 12): 401-2, 448-50, N, D '45; 65(1, 2):23-4, 81-3 Ja, F '46.* [284

Design, Size, and Validation of Sample for Market Research. American Marketing Association, Committee on Marketing Research. *J Marketing* 10(3):221-34 Ja '46.* [285

The Non-Central Wishart Distribution

Hypotheses. Ralph J. Brookner (Lt., USNR, Navy Board, Washington). *Ann Math Stat* 16(3):221–42 S '45.* [308

Industry's Use of Quality Control. A symposium edited by Martin A. Brumbaugh (Editor of *Industrial Quality Control* and Professor of Statistics, University of Buffalo). *Electrical Products:* Quality Control—A Tool of Management, pp. 4–8, by Joseph Manuele (Director of Quality Control, Westinghouse Electric Corp., East Pittsburgh, Pa.). *Radio Industry:* Peacetime Responsibilities and Opportunites in Quality Control, p. 8, by Andrew I. Peterson (Manager, Quality Control Department, Radio Corporation of America, Harrison, N. J.); Quality Control in the Radio Industry, p. 9, by J. Clark Ryan (Works Manager, Colonial Radio Corp.). *Steel Products:* Quality Control at Work, p. 9, by P. C. Clarke (Assistant General Manager, Hunter Pressed Steel Co., Lansdale, Pa.); The Economics of Quality Control at the English Needle & Fishing Tackle Co., Studley, Warwickshire, England, pp. 9–10, by William A. Bennett (Works Manager, Engineering Division). *Foundry:* Statistical Quality Control in the Foundry, p. 11, by R. W. S. Freeman (Ontario Research Foundation, Toronto, Canada). *Tractors:* Quality Control, pp. 11–2, by Frederick J. Halton (Assistant to the President, Deere and Co., Chicago, Ill.). *Gage Manufacture:* Quality Control-Management's Guardian, p. 12, by Clinton V. Johnson (Chief Engineer, Johnson Gage Co.); Quality Control a Bargain, pp. 12–3, by C. W. Kennedy (Manager, Quality Control, Federal Products Corp.). *Optical:* Post-War Expansion of Quality Control System, pp. 13–4, by Theodore B. Drescher (Vice-President in Charge of Manufacture, Bausch & Lomb Optical Co., Rochester, N. Y.). *Chemical:* Quality Control in Chemical Manufacture, p. 14, by James A. Mitchell (Supervisor of Quality Control, Cellulose Esters Division, Tennessee Eastman Corp.); Accuracy of Measurement, p. 15, by Maurice Millbourn. *Rubber:* Quality Control by Statistical Methods, p. 15, by G. H. Cassady (Manager, Quality Control Division, Fisk Tire Plant). *Textile:* Statistical Control a Permanent Feature, pp. 15–6, by A. G. Ashcroft (Director of Research, Alexander Smith & Sons Carpet Co., Yonkers, N. Y.). *Air Conditioning:* Quality Control, pp. 16–7, by E. F. Graumenz (Quality Engineer, Carrier Corp.). *Distribution:* Control Charts Locate Errors, p. 17, James M. Ballowe (Manager, Quality Inspection Department, Aldens Chicago Mail Order Co.). *Receiving Inspection:* Quality Control in Receiving Inspection, p. 17, by O. W. Hitchcock (Quality Manager, Scintilla Magneto Division, Bendix Aviation Corp.). *Sub-Assembly Inspection:* Sub-Assembly Inspection Pays Dividends, pp. 17–8, by H. H. Kent (Eastman Kodak Co.). *Final Inspection:* The Money Value of Industrial Statistics, p. 18, by Paul Peach (Industrial Statistician, Institute of Statistics, University of North Carolina, Raleigh). *Summary:* Implications of Statistical Quality Control, pp. 18–9, by Ralph E. Wareham (Consultant on Quality Control, New York, N. Y.). *Editor's Summary,* p. 19. *Ind Qual Control* 2(4A):4–19 Ja '46.* [309

A Report on Pre-War Quality. Martin A. Brumbaugh. *Ind Qual Control* 2(2):11–2 S '45.* [310

★**On the Statistical Treatment of the Results of Parallel Trials With Special Reference to Fishery Research.** H. J. Buchanan-Wollaston (Principal Naturalist on the staff of the Ministry of Agriculture and Fisheries, working at the Laboratories of the Freshwater Biological Association, Wray Castle, Ambleside, Westmorland, England). Scientific Publication No. 10. Ambleside, Westmorland, England: Freshwater Biological Association, Wray Castle, 1945. Pp. 65. Paper. 2s. 6d. *Reviewed in this issue.* [311

Statistical Methodology Index. No. 7. Oscar Krisen Buros (Associate Professor of Education, Rutgers University). *J Am Stat Assn* 40(232):439–52 D '45.* [312

Some Conclusions Regarding Resistance Welding and Statistical Quality Control. John Bayard Butler (Test Engineer, Hamilton Standard Propellers, East Hartford, Conn.). *Welding J* 24(10):982–4 O '45.* [313

What the Executive Should Know About Quality Control. Eugene Caldwell (Consulting Managing Engineer, Portland, Ore.). *Steel* 117(5):88+ Jl 30 '45.* [314

The Effect on a Distribution Function of Small Changes in the Population Function. Abstract. Burton H. Camp (Professor of Mathematics, Wesleyan University). *Ann Math Stat* 16(4):402 D '45.* [315

A Simple Orthogonal Multiple Factor Approximation Procedure. Hilding B. Carlson (Associate Professor of Psychology, Occidental College). *Psychometrika* 10(4):283–301 D '45.* [316

On Inductive Logic. Rudolf Carnap (Professor of Philosophy, University of Chicago). *Philos Sci* 12(2):72–97 Ap '45.* [317

Skeleton Tables Derived from the Experience of the Continuous Mortality Investigation. W. H. Chapman (Assistant

Actuary, Gresham Life Assurance Society, Ltd.). *J Inst Actuaries* 72(334):234–9 pt 2 '45.* [318

Distributions of Certain Frequency Constants in Samples From Non-Normal Populations. K. C. Cheriyan (Statistical Laboratory, Presidency College, Calcutta, India). *Sankhya* 7(2):159–66 N '45.* [319

Probability Theory: I, Background; II, Postulates of Experimental Method; III, Non-Mechanical Concepts. C. W. Churchman (Lecturer on Philosophy, University of Pennsylvania). *Philos Sci* 12(3):147–57, 158–64, 165–73 Jl '45.* [320

★**Statistical Methods in Quality Control.** P. C. Clarke (Assistant General Manager) and G. R. Armstrong (Quality Control Analyst). (Hunter Pressed Steel Co.) Lansdale, Pa.: Hunter Pressed Steel Co., 1945. Pp. 100 (pagination irregular). *Reviewed in this issue.* [321

The Use of Statistical Methods for Economic Control of Quality in Industry. Dudley J. Cowden (Professor of Economics, University of North Carolina) and William S. Connor (Statistician, Riverside and Dan River Cotton Mills, Danville, Va.). *South Econ J* 12(2):115–20 O '45.* [322

Probability, Frequency and Reasonable Expectation. R. T. Cox (Visiting Professor of Physics, Johns Hopkins University). *Am J Physics* 14(1):1–13 Ja–F '46.* [323

Nature, Nurture, and Probability. Letter. Gunnar Dahlberg (Head, State Institute of Human Genetics and Race Biology, Uppsala, Sweden). *Nature* 156(3966):530 N 3 '45.* [324

On the Validity of Statistical Tests of the Graduation of a Mortality Table. R. H. Daw. Abstract of discussion by: F. M. Redington, pp. 191–2; W. G. Bailey, pp. 192–3; L. E. Coward, pp. 193–4; H. W. Haycocks, p. 194; H. L. Seal, pp. 195–6; M. G. Kendall, pp. 196–8; M. Greenwood, p. 198; W. Perks, pp. 198–9, 201–2; H. E. Melville, pp. 199–200; R. H. Daw, pp. 200–1; K. A. C. Wheeler, p. 202. *J Inst Actuaries* 72 (334):174–90; discussion, 191–202 pt 2 '45.* [325

Number Defectives—Statistical Quality Control, pp. 9–21. D. J. Desmond (Research Chemist, Joseph Lucas Ltd., Great King St., Birmingham, England). In *Symposium of Papers on Statistical Quality Control*, 1945.* *See* 278. [326

The Probability Distribution Around a Fix in Celestial Navigation. Armin J. Deutsch (Yerkes Observatory, Williams Bay, Wis.). *Am J Physics* 13(6):379–83 D '45.* [327

A New Gaging Method for Quality Control. P. M. Dickerson (Lamp Division, Westinghouse Electric & Mfg. Co., Bloomfield, N. J.). *Tool & Die J* 10(12):132–4 Mr '45.* [328

Quality Control and Spot Welding: Application and Value of Statistical Methods. N. G. Dodd. *Aircraft Prod* 7 (80):503–4 D '45.* [329

Markoff Chains—Denumerable Case. J. L. Doob (Associate Professor of Mathematics, University of Illinois). *Trans Am Math Soc* 58(3):455–73 N '45.* Abstract: *B Am Math Soc* 51(11):880 N '45.* [330

Pearsonian Correlation Coefficients Associated With Least Squares Theory. Abstract. Paul S. Dwyer (Associate Professor of Mathematics, University of Michigan). *Ann Math Stat* 16(4):404 D '45.* [331

The Square Root Method and Its Use in Correlation and Regression. Paul S. Dwyer. *J Am Stat Assn* 40(232):493–503 D '45.* [332

Post-War Quality Control. George D. Edwards (Director of Quality Assurance, Bell Telephone Laboratories, Inc., New York). *Ind Qual Control* 2(4A):2–3 Ja '46.* [333

Operating Characteristics of Reject Limits for Measurements. Norbert L. Enrick (Quality Control Statistician, Inspection Service, U. S. Army Headquarters, New York). *Ind Qual Control* 2(2):9–10+ S '45.* [334

Tolerances on Assemblies. Benjamin Epstein (Staff Assistant, Quality Control Dept., Westinghouse Electric Corp., East Pittsburgh, Pa.). *Am Mach* 90(1):119–21 Ja 3 '46.* [335

A Technique for Analysis of Population Density Data. Ralph O. Erickson (Department of Botany, University of Rochester) and John R. Stehn (Research Division, Winchester Repeating Arms Co., New Haven, Conn.). *Am Midl Nat* 33(3):781–7 My '45.* [336

New Management Method for Industrial Processes: Reducing Defective Material and Inspection Time by Statistical Quality Control. H. H. Fairfield (Industrial and Metallurgical Engineer, Metallic Minerals Division, Bureau of Mines, Mines and Geology Branch, Department of Mines and Resources, Ottawa, Canada). *Can Metals & Met Ind* 8(8):32–6 Ag '45.* [337

The Fundamental Limit Theorems in Probability. W. Feller (Professor of Mathematics, Cornell University). *B Am Math Soc* 51(11):800–32 N '45.* [338

Note on the Law of Large Numbers and "Fair" Games. W. Feller. *Ann Math Stat* 16(3):301–4 S '45.* Abstract: *B Am Math Soc* 52(1):72 Ja '46.* [339

On the Normal Approximation to the

Binomial Distribution. W. Feller. *Ann Math Stat* 16(4):319-20 D '45.* Abstract: *B Am Math Soc* 52(1):72 Ja '46.* [340

On Statistical Problems in Test Construction. G. A. Ferguson (Cooperative Test Service, New York). *B Can Psychol Assn* 5(4):102-9 D '45.* [341

Life Quality Predictions by Statistical Methods. Enoch B. Ferrell (Bell Telephone Laboratories, Inc., New York). *Mach Design* 18(2):141-4 F '46.* [342

The Fractional Replication of Factorial Arrangements. D. J. Finney (Rothamsted Experimental Station, Harpenden, Herts, England). *Ann Eug* 12(4):291-301 N '45.* [343

Some Orthogonal Properties of the 4×4 and 6×6 Latin Squares. D. J. Finney. *Ann Eug* 12(4):213-9 N '45.* [344

The Logical Inversion of the Notion of the Random Sample. R. A. Fisher (Professor of Genetics, Cambridge University). *Sankhya* 7(2):129-32 N '45.* [345

A System of Confounding for Factors With More Than Two Alternatives, Giving Completely Orthogonal Cubes and Higher Powers. R. A. Fisher. *Ann Eug* 12(4):283-90 N '45.* [346

Sequential Sampling Inspection for Attributes. H. A. Freeman (Associate Professor of Statistics, Massachusetts Institute of Technology). *Ind Qual Control* 2(2):3-7 S '45.* [347

Lognormal Distributions. J. H. Gaddum (University of Edinburgh). *Nature* 156 (3964):463-6 O 20 '45.* [348

Further Remarks on Linkage Theory in Mendelian Heredity. Hilda Geiringer (Head, Department of Mathematics, Wheaton College). *Ann Math Stat* 16(4): 390-3 D '45.* [349

On the Definition of Distance in the Theory of the Gene. Hilda Geiringer. *Ann Math Stat* 16(4):393-8 D '45.* [350

On the Accuracy of the Different Approximations to the L_1-Distribution. Aleyamma George (University of Travancore, India). *Sankhya* 7(1):20-6 Ag '45.* [351

Using Statistics to Control Mass-Production Quality. Eugene Geddess (Head of Factory Engineering, North American Philips Co., Dobbs Ferry, N. Y.). *Factory Mgmt* 104(2):110-2 F '46.* [352

On the Distribution of the Estimate of Mean Deviation Obtained From Samples From a Normal Population. H. J. Godwin (Advisory Service on Statistical Method, Ministry of Supply, London). Appendix, Note on the Calculation of the Distribution of the Estimate of Mean Deviation in Normal Samples, pp. 257-8, by H. O. Hartley (Scientific Computing Service, Ltd., London). *Biometrika* 33(3):254-65 N '45.* [353

The Effect of a Number of Variables on a Quality Characteristic. Casper Goffman (Staff Assistant, Quality Control Dept., Westinghouse Electric Corp., East Pittsburgh, Pa.) *Ind Qual Control* 2(4):3-5 Ja '46.* [354

Composite Distributions. Abstract. Casper Goffman and Benjamin Epstein (Staff Assistant, Quality Control Dept., Westinghouse Electric Corp., East Pittsburgh, Pa.). *Ann Math Stat* 16(4):403 D '45.* [355

Systematic Sampling Pays in Controlling Filling Weights: Use in Can-Filling Weight Control: Shewhart Charts. H. P. Goode (Assistant Professor of Mechanical Engineering, Stanford University) and George I. Dundas (Sutter Packing Co., Palo Alto, Calif.). *Food Ind* 17(12):1468-70 D '45.* [356

Costs Dropped, Quality Bettered, Through Statistical Controls. Eugene L. Grant (Professor of Economics of Engineering, Stanford University). *Wings* [New York] 4(4):1716-9 S '45.* [357

Expected Values of a Statistical Variate. Abstract. E. J. Gumbel (Visiting Professor, New School for Social Research). *B Am Math Soc* 51(3):226 Mr '45.* [358

Population, Expected Values and Sample. Abstract. E. J. Gumbel. *Ann Math Stat* 16(4):403 D '45.* [359

Simplified Plotting of Statistical Observations. E. J. Gumbel. *Trans Am Geophys Union* 26(1):69-82 Ag '45.* [360

Research and Quality Control of Materials. E. H. Gurney (Chairman, Ontario Research Foundation, Toronto, Canada). *Tool Engineer* 14(12):42-6 Jl '45.* [361

A Basis for Analyzing Test-Retest Reliability. Louis Guttman (Associate Professor of Sociology, Cornell University). *Psychometrika* 10(4):255-82 D '45.* [362

Component Indexes as a Basis for Stratification in Sampling. Margaret Jarman Hagood and Eleanor H. Bernert. (Social Scientists, Bureau of Agricultural Economics, Washington.) *J Am Stat Assn* 40 (231):330-41 S '45.* [363

Moments of r and χ^2 for a Fourfold Table in the Absence of Association. J. B. S. Haldane (Professor of Biometry, University College, London). *Biometrika* 33(3):231-3 N '45.* [364

On a Method of Estimating Frequencies. J. B. S. Haldane. *Biometrika* 33(3):222-5 N '45.* [365

The Use of χ^2 as a Test of Homogeneity in a $(n\times2)$-Fold Table When Expectations Are Small. J. B. S. Haldane. *Biometrika* 33(3):234-8 N '45.* [366

The Theory of Unbiased Estimation. Abstract. P. R. Halmos (Assistant Professor of Mathematics, Syracuse University). *B Am Math Soc* 52(1):72–3 Ja '46.* [367

Limits for Control Charts. Preston C. Hammer (Partner, Probity Engineers, Los Angeles, Calif.) *Ind Qual Control* 2(3):9–11 N '45.* [368

On Optimum Estimates for Stratified Samples. Abstract. Morris H. Hansen (Statistical Assistant to the Director, Bureau of the Census, Washington) and William N. Hurwitz. *Ann Math Stat* 16(4): 403 D '45.* [369

On the Analysis of a Certain Six-by-Six Four-Group Lattice Design Using the Recovery of Inter-Block Information. Boyd Harshbarger (Professor of Statistics, Virginia Polytechnic Institute; and Statistician, Virginia Agricultural Experiment Station, Blacksburg, Va.). *Ann Math Stat* 16(4):387–90 D '45.* [370

Application of Quality Control to Resistance Welding. L. S. Hobson (Assistant Works Manager), R. S. Inglis (Superintendent of Quality Control and Inspection), and R. P. McCants (Welding Engineer, Works Laboratory). General Electric Co., Philadelphia, Pa.) *Elec Eng, Trans Sect* 64(8):573–5 Ag '45.* [371

Testing the Homogeneity of Poisson Frequencies. Paul G. Hoel (Assistant Professor of Mathematics, University of California, Los Angeles). *Ann Math Stat* 16(4):362–8 D '45.* [372

Quality Control: A Comparison of Compressed Limit Gauging Methods With Measurement Check in Dimensional Quality Control Practice. II. Howell. *Aircraft Prod* 7(85):530–43 N '45.* [373

Significance of Accurate Fine Measurements in the Statistical Prediction of Quality in Engineering Component Manufacture. Harry Howell. *Eng Insp* 10(1):4–16 sp '45.* [374

Statistical Quality Control: Some American Views on Its Fundamental Concepts. John M. Howell (Statistical Analyst, Manufacturing Control Department, Northrop Aircraft, Inc., Hawthorne, Calif.). *Aircraft Prod* 7(84):475–7 O '45.* [375

Some Combinatorial Formulas on Mathematical Expectation. L. C. Hsu (National Southwest Associated University, Kunming, China). *Ann Math Stat* 16(4):369–80 D '45.* [376

On the Power Functions of the E^2-Test and the T^2-Test. P. L. Hsu (Visiting Professor of Statistics, Columbia University). *Ann Math Stat* 16(3):278–86 S '45.* [377

On the Totality of the Solutions for the Symmetrical Incomplete Block Designs: $\lambda=2$, $k=5$ or 6. Q. M. Husain (Dacca University, Dacca, India). *Sankhyā* 7(2): 204–8 N '45.* [378

Statistical Methods in Anthropometric Studies in the Field of Nutritional Research. Rachel M. Jenss. *Am J Pub Health* 35(10): 1053–6 O '45.* [379

On the Average of a Certain Wiener Functional. Abstract. Mark Kac (Assistant Professor of Mathematics, Cornell University). *B Am Math Soc* 51(11):880 N '45.* [380

A Remark on Independence of Linear and Quadratic Forms Involving Independent Gaussian Variables. M. Kac *Ann Math Stat* 16(4):400–1 D '45.* [381

A Comment on the Correction of Reliability Coefficients for Restriction of Range. Hyman B. Kaitz (Social Security Board, Washington). *J Ed Psychol* 36(8):510–2 N '45.* [382

The Asymptotic Distribution of Runs of Consecutive Elements. Abstract. Irving Kaplansky (Assistant Professor of Mathematics, University of Chicago). *B Am Math Soc* 51(5):371 My '45.* [383

Multiple Matching and Runs by the Symbolic Method. Irving Kaplansky and John Riordon (Bell Telephone Laboratories, New York). *Ann Math Stat* 16(3):272–7 S '45.* [384

Scientific Procedure and Probability: A Symposium on Probability, Part II. Felix Kaufmann (New School for Social Research). *Philos & Phenom Res* 6(1):47–66 S '45.* [385

The Treatment of Ties in Ranking Problems. M. G. Kendall (Statistician, Chamber of Shipping of the United Kingdom, London, England). *Biometrika* 33(3):239–51 N '45.* [386

On the Design of Experiments for Weighing and Making Other Types of Measurements. K. Kishen (Department of Agriculture, Lucknow, India). *Ann Math Stat* 16(3):294–300 S '45.* [387

On the Design of Experiments for Weighing and Other Types of Measurements. Letter. K. Kishen. *Cur Sci* 14(8):194–5 Ag '45.* [388

Statistical Estimation of Simultaneous Economic Relations. Tjalling Koopmans (Research Associate, Cowles Commission, University of Chicago). *J Am Stat Assn* 40(232):448–66 D '45.* [389

A Note on Hotelling's T^2. Letter. P. V. Krishna Iyer (Imperial Agricultural Research Institute, New Delhi, India). *Cur Sci* 14(7):173–5 Jl '45.* [390

Tests of Significance by Analysis of Covariance in Multivariate Populations.

Letter. P. V. Krishna Iyer. *Cur Sci* 14(11): 207 N '45.* [391

The Use of Generalised Dirichlet's Integral in Solving Some Distribution Problems of Statistics. P. V. Krishna Iyer. *Proc Indian Acad Sci, Sect A* 22(2):75-83 Ag '45.* [392

Correction of G-M Counter Data. J. D. Kurbatov (Associate Professor of Physics) and H. B. Mann (Research Associate, Department of Mathematics). (Ohio State University.) *Physical R* 68(1 & 2):40-3 Jl 1 & 15 '45.* [393

Coefficient of Association Between Two Attributes in Statistics. M. Lakshmanamurti (Andhra University, Guntur, India). *Proc Indian Acad Sci, Sect A* 22(3):123-33 S '45.* [394

On the Use of Matrices in Certain Population Mathematics. P. H. Leslie (Bureau of Animal Population, Oxford University). *Biometrika* 33(3):183-212 N '45.* [395

Significance of Negative Results in Small Samples. Letter. Morton L. Levin and Hyman Goldstein (Division of Cancer Control, New York State Department of Health). *Sci* 102(2651):407 O 19 '45.* [396

The Origins of the Card System. G. J. Lidstone. *J Inst Actuaries* 72(334):229-33 pt 2 '45.* [397

Computational Technics. Irving Lorge (Executive Officer, Institute of Educational Research and Professor of Education, Teachers College, Columbia University). *R Ed Res* 15(5):441-6 D '45.* [398

Quality Control. E. Lott. *Eng Insp* 10(1): 30-9 sp '45.* [399

Note on the Validity of Mathematical **Probability in Sociological Research.** Thomas C. McCormick (Professor of Sociology and Anthropology, University of Wisconsin). *Am Sociol R* 10(5):626-31 O '45.* [400

Application and Organization of Statistical Quality Control in a Light Engineering Plant, pp. 22-37. D. S. McNaughton (Chief Inspector, Hercules Cycle and Motor Co., Minor Works, Birmingham 6, England). In *Symposium of Papers on Statistical Quality Control*, 1945.* *See* 278. [401

The Mode of Operation of Suppressant Variables. Quinn McNemar (Associate Professor of Psychology, Stanford University). *Am J Psychol* 58(4):554-5 O '45.* [402

Note on the Distribution of the Serial Correlation Coefficient. William G. Madow (Visiting Professor of Statistics, University of São Paulo, São Paulo, Brasil). *Ann Math Stat* 16(3):308-10 S '45.* [403

Report on the Bihar Crop Survey: Rabi Season 1943-44. P. C. Mahalanobis (Statistical Laboratory, Presidency College, Calcutta, India). *Sankhya* 7(1):29 106 Ag '45.* [404

Note on a Paper by C. W. Cotterman and L. H. Snyder. H. B. Mann (Research Associate, Department of Mathematics, Ohio State University). *Ann Math Stat* 16(3): 311-2 S '45.* [405

On the Frequency Theory of Probability: A Symposium on Probability, Part II. Henry Margenau (Professor of Physics, Yale University). *Philos & Phenom Res* 6(1):11-25 S '45.* [406

A Simple Test of Significance. Letter. A. Mateur. *Engineering* 160(4168):452 N 30 '45.* [407

Accounting and Statistics. Robert K. Mautz. *Acctg R* 20(4):399-410 O '45.* [408

A Simple Algebraic Development of Horst's Suppressor Variables. Paul E. Meehl (Instructor in Psychology, University of Minnesota). *Am J Psychol* 58(4):550-4 O '45.* [409

Statistics and Social Values. N. C. Mehta (Lucknow, India). *Sankhya* 7(2):167-72 N '45.* [410

Comments on Donald William's Paper: **A Symposium on Probability, Part II.** R. v. Mises (Professor of Aerodynamics and Applied Mathematics, Graduate School of Engineering, Harvard University). *Philos & Phenom Res* 6(1):45-6 S '45.* [411

A Method for Indexing Statistical Tabulations. I. M. Moriyama (Senior Social Science Analyst, Division of Vital Statistics, Bureau of the Census, Washington). *Estadistica* 3(11):391-4 S '45.* [412

High Strength Plate for the Navy. Abstract of 100. Charles M. Mottley (Lt. Com. USNR, Bureau of Ships, Washington). *Metal Prog* 48(4):750+ O '45.* [413

Statistics in the Service of Planning. Radhakamal Mukherjee (University of Lucknow, Lucknow, India). *Sankhya* 7(2): 173-5 N '45.* [414

Use of Automatic Calculation Methods in Aeronautical Research. Max M. Munk (Lecturer in Aerodynamics, Catholic University of America). *Aero Dig* 50(5):68 9+ S '45.* [415

Influence of Humidity and Temperature on the Yield of Cotton. K. Raghavan Nair and Purnendu Bose (Statistical Laboratory, Presidency College, Calcutta, India). *Sankhya* 7(2):213-6 N '45.* [416

A Simplified Method of Calculating Standard Errors. R. G. Newton. *Chem & Ind* (41):322-3 O 20 '45.* [417

Teaching and Research at the Statistical Laboratory, University of California. Jerzy Neyman (Director, Statistical Laboratory

and Professor of Mathematics, University of California, Berkeley). Biometrics B 1(6):83–4 D '45.* [418

★An Introduction to Educational Statistics. C. W. Odell (Associate Professor of Education, University of Illinois). New York 11: Prentice-Hall, Inc. (70 Fifth Ave.), 1946. Pp. xiii, 269. $3.50.* *To be reviewed.* [419

Statistical Quality Control of Methods and Materials. Edwin G. Olds (Associate Professor of Mathematics, Carnegie Institute of Technology). *Mat & Meth* 22(4): 1097–1101 O '45.* [420

Calculation of Markoff Chains by Incomplete Gamma and Beta Functions and by Charlier Polynomials. Abstract. Isaac Opatowski (Armour Research Foundation, Illinois Institute of Technology). *B Am Math Soc* 51(11):880–7 N '45.* [421

Direct and Reverse Transitions in Markoff Chains. Abstract. Isaac Opatowski. *B Am Math Soc* 51(9):665–6 S '45.* [422

Markoff Chains and Tchebychev Polynomials. Isaac Opatowski. *B Am Math Soc* 52(1):73 Ja '46.* [423

Fraction Defective of Battery Adapter Used in Handie-Talkie. Frank A. Palumbo and Edward S. Strugala (Assistant Production Engineer). (Galvin Mfg. Corp., Chicago, Ill.) *Ind Qual Control* 2(3):6–8+ N '45.* [424

★An Introduction to Industrial Statistics and Quality Control. Paul Peach (Industrial Statistician, Institute of Statistics, University of North Carolina, Raleigh). Raleigh, N. C.: L. B. Phillips Co., 1945. Pp. ix, 150. Out of print. Paper. *Reviewed in this issue.* [425

The Probability Integral of the Mean Deviation: Editorial Note. E. S. Pearson (Department of Statistics, University College, London). *Biometrika* 33(3):252–3 N '45.* [426

Research Methods and Designs. Charles C. Peters (Director of Educational Research, Pennsylvania State College), Agatha Townsend (Educational Records Bureau, New York), and Arthur E. Traxler (Associate Director, Educational Records Bureau). *R Ed Res* 15(5):377–93 D '45.* [427

Fallacious Arguments From Experiments on Methods of Teaching. S. J. F. Philpott. *Brit J Ed Psychol* 15(2):57–69 Je '45.* [428

Generalisation of Markoff's Theorem and Tests of Linear Hypotheses. C. Radhakrishna Rao (Statistical Laboratory, Presidency College, Calcutta, India). *Sankhya* 7(1):9–16 Ag '45.* [429

Markoff's Theorem With Linear Restrictions on Parameters. C. Radhakrishna Rao. *Sankhya* 7(1):16–9 Ag '45.* [430

Statistical Quality Control: What It Is and What It Does. William B. Rice (Director, Department of Statistics and Reports, Plomb Tool Co., Los Angeles, Calif.). *Ind Qual Control* 2(4):6–10 Ja '46.* [431

Some Applications of Statistical Methods to Fishery Problems. William E. Ricker (Associate Professor of Zoology, Indiana University). *Biometrics B* 1(6):73–9 D '45.* [432

On the Measure of a Random Set, II. H. E. Robbins (Lt., USNR, Postgraduate School, U. S. Naval Academy). *Ann Math Stat* 16(4):342–7 D '45.* [433

On the Variance of a Random Set in *n* Dimensions. Abstract. H. E. Robbins. *Ann Math Stat* 16(4):402 D '45.* [434

Misuse of the Logarithmic Scale in Charting Business Trends. Leslie F. Robbins (Purchasing Agent, University of Colorado). *Purchasing* 18(4):82–4 Ap '45.* [435

How Quality Control Can Improve Batch Production. W. J. Roberts. *Ind Power & Prod* 21(232):47–8+ F '45.* [436

Associated Statistical Techniques, pp. 67–74. J. W. Rodgers (Chief Metallurgist, English Needle and Fishing Tackle Co. Ltd., Studley, Warwickshire, England). In *Symposium of Papers on Statistical Quality Control*, 1045.* See 278. [437

Practical Application of Statistical Methods in a Quality Control Program. W. T. Rogers (Metallurgical Statistician, National Tube Co., Lorain, Ohio). *Blast F & Steel Pl* 34(2):233–5 F '46.* [438

The Individual Sampling Distribution of the Maximum, the Minimum and Any Intermediate of the *p*-Statistics on the Null-Hypothesis. Samarendra Nath Roy (Statistical Laboratory, Presidency College, Calcutta, India). *Sankhya* 7(2):133–58 N '45.* [439

Bernoulli's Theorem and Tchebychef's Analogue. Samarendra Nath Roy and Purnendu Bose (Statistical Laboratory, Presidency College, Calcutta, India). *Sankhya* 7(2):209–10 N '45.* [440

Comment on Post-War Quality Control of Commercial Products. Letter. R. Royan. *Eng Insp* 10(1):30 sp '45.* [441

Errors in the Routine Daily Measurement of the Puerperal Uterus. C. Scott Russell (First Assistant, Nuffield Department of Obstetrics and Gynaecology, University of Oxford). *Biometrika* 33(3): 213–21 N '45.* [442

The U. S. Army Ordnance Department's

Actuaries. Irregularly issued in recent years; (6s.); Cambridge University Press, Bentley House, London.

J Marketing—The Journal of Marketing. 4 issues; $4; 450 Ahnaip St., Menasha, Wis.

J Res Nat Bur Stand—Journal of Research of the National Bureau of Standards. 12 issues in 2 vols.; $3.50(30¢); U. S. Government Printing Office, Washington 25, D. C.

Mat & Meth—Materials and Methods. Formerly *Metals and Alloys.* 12 issues; $2 (25¢); Reinhold Publishing Corp., 330 West 42nd St., New York 18, N. Y.

Math Student—The Mathematics Student. Published irregularly in recent years; G. A. Srinivasan, Presidency College, Madras, India.

Mech World—Mechanical World. 52 issues; 25s.(6d.); Emmott & Co. Ltd., 31 King St. West, Manchester 3, England.

Metal Prog—Metal Progress. 12 issues; $5($1); American Society for Metals, 7301 Euclid Ave., Cleveland, Ohio.

Mfg & Ind Eng—Manufacturing and Industrial Engineering. 12 issues; Consolidated Press Ltd., 73 Richmond St. West, Toronto 1, Ont., Canada.

Mill & Factory. 12 issues; Conover-Mast Corp., 205 East 42nd St., New York 17, N. Y.

Monetary Times—The Monetary Times. 13 issues; $3(25¢); 341 Church St., Toronto 2, Canada.

Nat Underw—National Underwriter. 53 issues; $5(20¢); 175 W. Jackson Blvd., Chicago, Ill.

Philos Sci—Philosophy of Science. 4 issues; $5($1.25); Mount Royal & Guilford Avenues, Baltimore 2, Md.

Physical R—The Physical Review. 6 issues in 2 vols.; $16.50($1.50); American Institute of Physics, 57 East 55th St., New York 22, N. Y.

Proc Indian Acad Sci, Sect A—Proceedings of the Indian Academy of Sciences, Section A. 4 issues; Rs. 18 (Rs. 2 or 3s.); Indian Academy of Sciences, Bangalore, India.

Purchasing. 12 issues; $3(35¢); Conover-Mast Co., 205 East 42nd St., New York 17, N. Y.

R Ed Res—Review of Educational Research. 5 issues; $4($1); 1201 16th St., N. W., Washington 6, D. C.

Rock Products. 12 issues; $2(25¢); Maclean-Hunter Publishing Corp., 309 West Jackson Blvd., Chicago 6, Ill.

Sankhya—Sankhya: The Indian Journal of Statistics. 4 issues; Rs. 30(Rs. 10); Presidency College, Calcutta, India.

Sci—Science. 52 issues in 2 vols.; $6(15¢); Permanent Secretary, A.A.A.S., Smithsonian Institution Bldg., Washington 25, D. C.

South Econ J—The Southern Economic Journal. 4 issues; $3($1); Box 190, Chapel Hill, N. C.

Timberman—The Timberman. 12 issues; $3(25¢); 519 S.W. Park Ave., Portland 5, Ore.

Tool & Die J—The Tool and Die Journal. 12 issues; 2490 Fairmount Blvd., Cleveland 6, Ohio.

Tool Engineer—The Tool Engineer. 12 issues; $6; 550 W. Lafayette Blvd., Detroit 26, Mich.

Trans Am Geophys Union—Transactions of the American Geophysical Union. 6 issues; $7.50($1.50); 5241 Broad Branch Road, Washington 15, D. C.

Trans Am Math Soc—Transactions of the American Mathematical Society. 6 issues in 2 vols.; $12; 531 West 116th St., New York 27, N. Y.

Wings [New York]. 12 issues; McGraw-Hill Publishing Co., 330 West 42nd St., New York 18, N. Y.

// JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 234 JUNE 1946 *Volume 41*

THE MEMBERSHIP OF THE AMERICAN STATISTICAL ASSOCIATION—AN ANALYSIS

ABNER HURWITZ AND FLOYD C. MANN
Bureau of Labor Statistics

IF MR. "X" were a composite of the ASA membership, he would most likely be living in Washington, D. C. or New York City, employed by a government agency as an economist or statistician specializing either in planning and reviewing surveys or collecting or directing the collection of statistical data. His principal field of interest in statistical data would center around economics and economic theory, and if he belonged to another professional society it would probably be the American Economics Association He would have received an advanced degree in one of the social sciences or its disciplines, and would at some time have published an article or a book.

These are the predominant characteristics of the members of the American Statistical Association as reported on a special directory form, which was distributed to all members who could be reached during the early part of 1945, and the information furnished in support of membership applications by those persons joining the Association between January and September 1945. The 1945 questionnaire requested more information about each member than had been obtained in earlier directory forms, and this article, therefore, presents a more comprehensive tabular analysis[1] of the characteristics and interests of the membership than has been available previously.[2] Data were obtained on present occupation and organization by which employed, positions

[1] The detailed tabular analyses presented in this article were made possible through the cooperation of Joseph F. Drager, Chief of the Machine Tabulation Division of the Bureau of Labor Statistics.

[2] For earlier analyses of the composition of the membership of the American Statistical Association, see Willford I. King: "Classification of Members of American Statistical Association on Basis of Duties and Interests," this JOURNAL, Vol. 22, June 1927, pp. 224–226; Stuart A. Rice and Morris Green: "Interlocking Memberships of Social Science Societies," this JOURNAL, Vol. 24, September 1929, pp. 303–306, and "Composition of the American Statistical Association," this JOURNAL, Vol. 25, June 1930, pp. 198–202; and Richard L. Funkhouser: "Membership of the American Statistical Association on its Hundredth Anniversary," this JOURNAL, Vol. 36, September 1941, pp. 329–341.

held since 1940, education, major and minor statistical specialties, fields of interest in statistical data and methods, membership in other professional societies, and chapter and national offices held. Additional occupational data were reported by those members on leave from a company, university, or other organization.

More than 3300 persons were listed in the 1945 Membership Directory but the largest number of persons reporting information for any one question was 2942. The following tables are based on the total number of members furnishing information in answer to each question.

DISTRIBUTION OF MEMBERSHIP BY TYPES OF EMPLOYING ORGANIZATIONS

Almost 35 per cent of the members reporting their employing organization in 1945 were employed by government agencies, one-sixth were with colleges and universities, and one in ten was associated with a manufacturing or industrial company. Twelve per cent were in the military services. Table 1 shows the types of organizations with which

TABLE 1

TYPE OF ORGANIZATION BY WHICH MEMBERS OF THE AMERICAN STATISTICAL ASSOCIATION WERE EMPLOYED IN 1940 AND 1945

Type of Organization	Members Reporting Employment			
	1945			1940*
	Number	Per Cent	Per Cent Excluding Military	Per Cent
All A.S.A. members reporting:	2942	100.0	100.0	100.0
Government agencies:	1019	34.6	39.4	38.0
Federal	862	29.3	33.3	29.1
State	130	4.4	5.0	7.3
Local	27	0.9	1.1	1.6
Colleges and universities	492	16.7	19.0	21.4
Financial institutions	202	6.9	7.8	10.4
Public utilities, transportation, or communications companies	115	3.9	4.5	3.0
Manufacturing industries	324	11.0	12.5	5.2
Retail and wholesale trades	21	0.7	0.8	1.5
Other organizations:	373	12.7	14.4	16.0
Business research	143	4.9		
Trade associations	62	2.1		
Research foundations	60	2.0		
Charitable organizations	36	1.2		
Other	72	2.5		
Military services	354	12.0		
Retired	26	0.9	1.0	1.6
None	16	0.6	0.6	. .

* Source: Richard L. Funkhouser, *op. cit.*, p. 330.

members reported employment in 1945 and provides a comparison with the membership's employment connections in 1940.[3]

A comparison of the 1940 and 1945 organizational distributions provides some evidence of how employment of the membership of the American Statistical Association was affected by World War II. In spite of the withdrawal of members into the military services, the per cent of members employed by the Federal Government, public utility, transportation and communications companies remained unchanged and the per cent of members in manufacturing industries increased two-fold. The number of members connected with colleges and universities, financial institutions, trades, and other organizations decreased considerably.

This resulted from members shifting into government and industry from universities and other non-war activities, and the attraction of new members into the Association from the greatly increased Federal payroll and from industrial plants in which the widespread introduction of quality control techniques attracted many users of statistical methods.[4] These comparisons are more striking when percentages for 1945 are computed excluding members in military services. Of the civilian members who reported the organizations by which they were employed, one-third were working in Federal agencies and one-eighth were employed in manufacturing.

[3] The above distributions are not strictly comparable with that prepared under the direction of W. I. King in 1927. However, certain groupings can be made to provide a rough measure of the per cent of members employed by different types of organizations since 1926.

Combinations of Organizations	Per Cent		
	1926	1940	1945
Government agencies, colleges and universities (including federal, state and municipally employed)	42.8	61.4	51.3
Financial institutions (including insurance)	15.0	10.4	6.0
Public service corporations	8.4	3.9	3.9
Manufacturing (including publishers and printing)	12.6	5.2	11.0
Mercantile	3.0	1.5	0.7
Professional service organizations (including business research, charitable and other organizations)	8.4	16.0	8.6
Trade associations	4.1	—	2.1
Foundations, etc.	5.7	—	2.0
Retired	—	1.6	0.9
Military	—	—	12.0
None	—	—	0.6
Total	100.0	100.0	100.0

[4] The increasing use of statistical controls in industry is evidenced by the organizational affiliations of the more than 400 members elected to the Association between September 1945 and May 1946. More than 25 per cent of these new members are employed in manufacturing plants as statisticians, quality control engineers, chemists and metallurgists, and another 11 per cent are connected with industry in other capacities.

OCCUPATIONAL DISTRIBUTION OF MEMBERSHIP

Nearly 29 per cent of the members reporting occupational information identified themselves as economists, 27 per cent had the rank or title of statistician, and almost one-eighth were executives, directors, or managers. The remaining third reported a wide range of occupations,

TABLE 2

OCCUPATIONS OF MEMBERS OF THE AMERICAN STATISTICAL ASSOCIATION IN 1915

Occupation	Members Reporting Occupation			
	Number		Per Cent	
All A.S.A. members reporting:		2550		100.0
Economists in the fields of:		730		28.6
Prices, production, etc.	117		4.6	
Labor and employment	84		3.3	
Finances, taxes, and insurance	93		3.6	
Marketing and trades	96		3.8	
Agriculture	46		1.8	
Transportation, communication, and public utilities	27		1.0	
Social science	24		1.0	
Other research	111		4.3	
Teaching	132		5.2	
Statisticians in the field of:		684		26.8
Economics	172		6.7	
Social sciences	148		5.8	
Physical and biological sciences	22		0.9	
Mathematics	27		1.1	
Business and management	62		2.4	
Industry	70		2.7	
Other research	107		4.2	
Teaching	78		3.0	
Executives, directors, and managers		304		11.9
Other administrative officers		49		1.9
Auditors and accountants		69		2.7
Marketing specialists		87		3.4
Mathematicians and actuaries		101		4.0
Engineers, industrial		28		1.1
Engineers, other		25		1.0
Psychologists		48		1.9
Sociologists		38		1.5
Social scientists		35		1.4
Physicians and public health officers		33		1.3
Journalists		24		1.0
Chemists		24		1.0
Biologists		18		0.7
Botanists		19		0.7
Other professions		88		3.4
Research directors		31		1.2
Research assistants		50		2.0
Clerical		46		1.8
Students		19		0.7

including mathematicians, marketing specialists, sociologists, and physical scientists. The heavy occupational representation of the social science fields in the Association is shown in Table 2.

The information in the above table was obtained from the first section of the 1945 Directory questionnaire, which asked for data on each member's "rank or title" and "description of duties or functions." The problems encountered in attempting to classify this type of data into mutually exclusive occupational categories were formidable, and the classifications used provide only a rough index to functions and job responsibilities of the members. Persons who had titles as chiefs of government agency units but whose duties, according to the descriptions provided, were primarily those of an economist, social scientist, etc., were classified according to the description. No person was classified as a statistician unless his job title or the description of his duties indicated definitely that he should be so classified. War-time security restrictions precluded many members in the military services from furnishing specific job descriptions. These members were, therefore, not included in this summary.

Since more than half of the Association's members identified themselves as statisticians and economists, these two groupings were classified further according to the principal field in which the member's job responsibility was centered. Forty-six per cent of the "statisticians" were working in economics or one of the other social science fields; 18 per cent of the economists were teaching, one in six was working in the fields of prices and production, and 13 per cent in marketing or the trades.

In addition to the more than 8 per cent of the membership who were teaching economics and statistics, another 8 per cent were engaged in teaching in other fields, the most important of which were mathematics, business administration, and psychology. Members teaching subjects other than economics and statistics are not shown separately in Table 2, but were included under their major occupational group. Students in September 1945 represented less than one per cent of the membership.[5]

OCCUPATIONAL DISTRIBUTION BY EMPLOYING ORGANIZATION

Table 3 shows the occupations of the members reporting this information by the type of organization in which they were employed in 1945. More than one-half of the statisticians in the Association were engaged in government work, 14 per cent were employed by universities

[5] Between September 1945 and May 1946 the number of student members more than doubled.

TABLE 3

OCCUPATIONS OF AMERICAN STATISTICAL ASSOCIATION MEMBERS BY EMPLOYING ORGANIZATION IN 1945

Occupations in 1945	Members Reporting Employment in										Members in Military Services in 1945[*]	
	All Organizations		Government Agencies	Universities and Colleges	Financial Institutions	Public Utilities, etc.	Manufacturing and Trades	Business Research and Trade Ass'ns.	Other Organizations	Organization not Reported		
	Number	Per Cent									Number	Per Cent
All A.S.A. members reporting—Number	2586	—	1019	492	202	115	345	205	168	40	354	100.0
Per cent	100.0	—	39.4	19.0	7.8	4.5	13.3	7.9	6.5	1.6	—	—
			(Per cent of members reporting)									
Statisticians	684	100.0	51.8	14.3	3.2	6.0	12.1	5.7	6.3	0.6	71	20.0
Economists	730	100.0	54.7	19.4	5.9	3.0	6.0	5.9	4.0	1.1	78	22.0
Executives, directors and managers	304	100.0	17.7	6.6	26.0	5.9	19.4	14.5	6.6	3.3	20	5.7
Other administrative officers	49	100.0	44.9	0.0	4.1	10.2	26.5	4.1	10.2	0.0	7	2.0
Accountants and auditors	69	100.0	18.8	2.9	13.0	11.6	34.9	11.6	5.8	1.4	6	1.7
Marketing specialists	87	100.0	2.3	1.1	18.4	2.3	33.4	40.2	2.3	0.0	10	2.8
Engineers (including quality control engineers)	53	100.0	5.7	7.5	0.0	15.1	60.4	7.5	3.8	0.0	7	2.0
Mathematicians and actuaries	101	100.0	17.8	53.4	19.8	0.0	2.0	4.0	0.0	3.0	8	2.3
Psychologists, sociologists and social scientists	121	100.0	32.3	45.4	0.0	0.0	3.3	3.3	14.9	0.8	17	4.7
Biologists, botanists and chemists	61	100.0	27.9	27.9	0.0	0.0	29.4	3.3	11.5	0.0	3	0.8
Research directors and assistants	81	100.0	32.2	14.8	4.9	1.2	11.1	11.1	24.7	0.0	7	2.0
Other professionals	145	100.0	20.7	45.5	0.7	1.4	9.0	4.1	9.6	9.0	14	4.0
Clerks	46	100.0	49.9	2.2	8.7	8.7	10.9	10.9	8.7	0.0	9	2.6
Students	19	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	2	0.6
Occupations not reported	36	100.0	52.9	2.8	5.5	11.0	27.8	0.0	0.0	0.0	95	26.8
Members in military services in 1945[*]												
Number	354	—	136	44	16	14	27	1	26	90	—	—
Per cent	100.0	—	35.4	12.4	4.5	4.0	7.6	0.3	7.4	25.4	—	—

[*] Occupations and organizations in which members were employed before entering services.

and colleges, chiefly as teachers, and 12 per cent by manufacturing and business organizations. Fifty-five per cent of the economists were with the government and almost half of these were working in the fields of prices, production, labor, and employment. About 75 per cent of the executives, directors, and managers were connected with business organizations.

One-fourth of the members in the military in 1945 did not report their occupations or the type of organization with which they were associated prior to entering the services. However, more than half of those reporting had been employed as government statisticians and economists. Sixteen per cent had been connected with universities and colleges.

TYPES OF STATISTICAL WORK PERFORMED

Slightly more than 50 per cent of the members of the American Statistical Association indicated that they had special skills in the planning and review of plans for surveys and in the collection or the direction of the collection of statistical data. About 48 per cent considered themselves to be skillful users of statistical data and to be proficient in sampling and analyzing data. Nearly twice as many indicated special abilities for general research as for teaching.

Although more than half of the membership specialized in planning and reviewing of plans for surveys, less than one out of every five considered this to be his major specialty. Sampling and the analysis of statistical data was checked by 15 per cent as a major skill.

The relationship between the types of statistical work regarded by the membership as major specialties and the types checked as other special skills is shown in Table 4. The different types of statistical work have been ranked according to the frequency with which each was checked either as a major or an "other" special skill.

On the average the different types of statistical work were checked as "other special skills" 2.6 times as frequently as they were indicated as "major specialties." It is not surprising that the more specialized types of work—classification; graphic presentation; index number construction; editing, coding, or tabulating—were less frequently designated as major specialties and more frequently checked as "other special skills." More than 7 times as many considered graphic presentation as a minor skill as compared with the number of members who designated it as their major specialty.

Space does not permit the presentation of an analysis in this article of the major patterns of statistical work in which the members reported

TABLE 4

SPECIAL SKILLS IN STATISTICAL WORK PERFORMED BY AMERICAN STATISTICAL ASSOCIATION MEMBERS IN 1945*

Type of Statistical Work	Members Reporting Statistical Work Performed as					
	Total		Major Specialty		Other Special Skill	
	Number	Per Cent	Number	Per Cent	Number	Per Cent
All A.S.A. members reporting:	2715	100.0	—	—	—	—
Planning and review of plans for surveys	1372	50.5	510	18.8	862	31.7
Collecting or directing the collection of data	1368	50.4	355	13.1	1013	37.3
Use of statistical data	1304	48.0	323	11.9	981	36.1
Sampling and analysis	1300	47.9	416	15.3	884	32.6
General research	1189	43.8	351	12.9	838	30.9
Editing, coding, and tabulating	850	31.3	154	5.7	696	25.6
Graphic presentation	809	29.8	95	3.5	714	26.3
Teaching	622	22.9	266	9.8	356	13.1
Index number construction	346	12.7	50	1.8	296	10.9
Quality control	309	11.4	74	2.7	235	8.7
Classification	182	6.7	15	0.6	167	6.1
Actuarial methods	151	5.6	63	2.3	88	3.3
Other statistical work	32	1.2	4	0.2	28	1.0

* 528 members who reported type of statistical work performed did not indicate any skill as their major specialty. 436 members indicated two or more types of work as their major skill, and thus there is duplication in these figures.

specialization. However, survey planning, data collection, sampling, and editing, coding and tabulation appear most often in combination.

FIELDS OF INTEREST IN STATISTICAL DATA AND METHODS

A larger number of the members indicated interests in statistical data and methods in the fields of economics and economic theory than in any other fields. Prices and production was the subject most frequently designated as of primary importance. One-third of the Association checked this field as an area in which their interests centered. One in four reported interest in labor and employment economics, 20 per cent noted population research as an interest, and more than one-seventh rated the following as areas of primary importance: domestic and international trade, marketing and advertising, sociology and social service, vital statistics and public health, and mathematical sciences.

As would be expected, the number of members interested in different fields varies considerably when these areas of concentration are classified as of primary, secondary, or minor importance. While 12 per cent indicated population research to be a field in which their interest in

statistical data and methods centered, less than 3 per cent ranked it as an interest of primary importance. Similarly, although only one per cent rated opinion polling as their first interest, 8 per cent reported it as an interest field.

The first six fields which the membership ranked as of primary importance were: prices and production, labor and employment, general economics and theory, marketing and advertising, vital statistics and

TABLE 5

FIELDS OF INTEREST IN STATISTICAL DATA AND METHODS OF MEMBERS OF THE AMERICAN STATISTICAL ASSOCIATION IN 1945

Fields of Interest in Statistical Data and Methods	All Members Reporting Interest		Members Ranking Interests as of: Primary Importance		Secondary Importance		Minor Importance	
	Number	Per Cent	Number	Per Cent	Number	Per Cent	Number	Per Cent
All A.S.A. members reporting:	*2706	100.0	—	—	—	—	—	—
Economics and economic theory:								
General interest	460	16.6	182	6.6	213	7.7	65	2.3
Prices and production	934	33.8	384	13.9	336	12.2	214	7.7
Labor and employment	650	23.5	269	7.6	173	6.2	208	9.7
Domestic and international trade	415	15.0	55	2.0	148	5.3	212	7.7
Public finance and taxation	354	12.8	78	2.8	91	3.3	185	6.7
Transportation and public utilities	244	8.8	86	3.1	57	2.1	101	3.6
Business and management:								
General interest	338	12.2	109	3.9	100	3.6	129	4.7
Marketing and advertising	421	15.2	134	4.8	126	4.6	161	5.8
Accounting and private finance	339	12.3	108	3.9	100	3.6	131	4.8
Personnel	268	9.7	34	1.2	88	3.2	146	5.3
Production design and prod. control	229	8.3	62	2.2	63	2.3	104	3.8
Population research	547	19.8	69	2.5	148	5.3	331	12.0
Sociology and social service	415	15.0	125	4.5	84	3.0	206	7.5
Vital statistics and public health	408	14.8	126	4.6	91	3.3	191	6.9
Mathematical sciences	382	13.8	104	3.8	84	3.0	194	7.0
Opinion polling	311	11.2	34	1.2	67	2.4	210	7.6
Agricultural economics and rural statistics	297	10.7	89	3.2	53	1.9	155	5.6
Psychology	252	9.1	60	2.2	52	1.9	140	5.0
Education	243	8.8	56	2.0	60	2.2	127	4.6
Biological sciences	237	8.6	59	2.1	68	2.5	110	4.0
Physical sciences and engineering	200	7.3	54	2.0	54	2.0	92	3.3
Medicine	199	7.2	41	1.5	82	3.0	76	2.7
Agricultural experimentation and production	143	5.2	41	1.5	38	1.4	64	2.3
Chemical sciences	75	2.7	10	0.4	18	0.6	47	1.7
Other fields of interest	61	2.2	38	1.4	11	0.4	12	0.4

* 420 members reported but did not indicate the order of importance of their interests. The primary interests of these members were determined by observation of related information (e.g. occupation, major field of education) but ranked as secondary since no major importance was actually indicated.

public health, and sociology and social service. Table 5 shows how the members rated their different fields of interest in statistical data and methods according to importance.

Of those members ranking prices and production as their field of chief interest, 23 per cent indicated that domestic and international trade was their second area of interest, and about the same number checked labor and employment as of minor importance. The predominant interest pattern of those ranking marketing and advertising as primarily important showed prices and production and opinion polling as second and third interests, respectively. Other predominant patterns of interests are as follows:

Product design and production control, prices and production, and physical sciences and engineering.
Sociology and social service, population research and census, and vital statistics and public health.
Education and psychology.
Agricultural experimentation and biological sciences.
Physical sciences, engineering, and mathematical sciences.

AFFILIATION OF MEMBERSHIP WITH OTHER PROFESSIONAL SOCIETIES

Almost 34 per cent of the American Statistical Association members answering the question concerning memberships in other professional societies indicated that they had no other affiliations, 28 per cent belonged to one other society, 17 per cent were affiliated with two, and the remaining 21 per cent held memberships in three or more other professional organizations.

Two out of every five members who were affiliated with other groups belonged to the American Economic Association; almost 17 per cent were members of the Institute of Mathematical Statistics. In Table 7 the societies which were listed on the directory form have been grouped according to subject matter; the unlisted societies entered on the directory form by members have also been included in these general categories. Since 38 per cent of the membership reporting indicated that they belonged to two or more other societies, extensive duplications exist in the first two columns of this table.

More than one-half of the members belonging to at least one other society indicated that they were affiliated with an economics group, more than one-third were affiliated with one or more of the "political and social science" societies, and almost 23 per cent were members of other associations. More than one out of every five members be-

longed to a mathematical or other statistical organization. Eleven per cent were associated with one of the societies classified under business management, and six per cent were members of psychological and educational associations. Although the members of the American Statistical Association were for the most part interested in the social science fields, it is notable that almost 13 per cent of the membership belonging to other professional societies were members of physical and biological science associations, and almost 15 per cent belong to the American Association for the Advancement of Science, an organization composed of persons having a diversity of interests in the sciences.

Some measure of the diversity of interests of the individual American

TABLE 6

NUMBER OF OTHER PROFESSIONAL SOCIETIES TO WHICH AMERICAN STATISTICAL ASSOCIATION MEMBERS BELONGED IN 1945

Number of Professional Societies with which Affiliated	A.S.A. Members	
	Number	Per Cent
All A.S.A. members reporting:	2796	100.0
None other	942	33.7
One or more other	1854	66.3
One other	787	28.2
Two others	479	17.1
Three others	306	10.9
Four others	148	5.3
Five others	60	2.1
Six others	45	1.6
Seven others	12	0.4
Eight others	7	0.3
Nine or more others	11	0.4

Statistical Association members is provided by the data in Table 7 which show those members who belonged to one other society or to only societies classified under one major subject matter group. Fifty-eight per cent of the members who belonged to other societies were associated with only one of these major groups, about 27 per cent belonged to one or more economics associations, and 12 per cent held memberships in "political and social science" societies. Although more than 22 per cent of the American Statistical Association members who belonged to other societies were members of some mathematical or other statistical association, only 10 per cent belonged exclusively to societies in this group.

Diversity of interests of individual American Statistical Association members is apparent from figures which show that almost 42 per cent

TABLE 7

AMERICAN STATISTICAL ASSOCIATION MEMBERS' AFFILIATIONS WITH OTHER PROFESSIONAL SOCIETIES IN 1955

Professional Societies and Groups	A. S. A. Members					
	Belonging to One or More Other Professional Societies		Belonging to Only One Other Professional Society		Belonging to Societies in Only One Group*	
	Number	*Per Cent*	*Number*	*Per Cent*	*Number*	*Per Cent*
All A. S. A. members reporting affiliation with other societies:	1851	100.0	787	42.5	1082	58.4
Economic associations:	965	52.0	392	21.1	505	27.3
American Economic Association (1)	720	39.3	270	14.6		
American Marketing Association (4)	215	11.6	73	3.9		
Econometric Society (6)	173	9.3	13	0.7		
American Farm Economic Association (12)	76	4.1	22	1.2		
Other economic associations	102	5.5	14	0.8		
Political and social science associations:	643	34.7	158	8.5	222	12.0
American Academy of Political & Social Science (5)	196	10.6	22	1.2		
Academy of Political Science (7)	152	8.2	23	1.2		
American Public Health Association (8)	149	8.0	48	2.6		
Population Association of America (9)	134	7.2	12	0.6		
American Sociological Society (10)	102	5.5	15	0.8		
American Political Science Association	38	2.0	4	0.2		
Rural Sociological Society	23	1.2	1	0.1		
Other political and social science associations	187	10.1	33	1.8		
Mathematical and statistical associations:	417	22.5	107	5.8	182	9.8
Institute of Mathematical Statistics (2)	313	16.9	70	3.8		
Actuarial Society of America	35	1.9	1	0.1		
Institute of Actuaries	34	1.8	2	0.1		
National Society of Quality Control Engineers	33	1.8	7	0.4		
Other mathematical and statistical associations	203	10.9	27	1.5		
Business management associations:	210	11.3	49	2.6	56	3.0
Society for the Advancement of Management (14)	50	2.7	10	0.5		
American Management Association	43	2.3	6	0.3		
National Association of Cost Accountants	40	2.2	15	0.8		
American Society for Public Administration	34	1.8	3	0.2		
Other business management associations	54	2.9	15	0.8		
Psychological and educational associations:	112	6.0	21	1.1	29	1.6
American Psychological Association (11)	77	4.2	14	0.8		
Psychometric Society (13)	57	3.1	4	0.2		
American Educational Research Association	36	1.9	1	0.1		
Other psychological and educational ass'ns.	13	0.7	2	0.1		
Other associations	421	22.7	60	3.2	89	4.8
American Ass'n. for the Advancement of Science (3)	272	14.7	16	0.9		
Physical and Biological Associations	241	13.0	44	2.4		

* 773 (41.8%) reporting members belonged to associations classified in two or more groups; 556 (29.9%) members were associated with two groups; 180 (9.7%) with three groups; and 37 (2.0%) with four or five groups.

of those members who were affiliated with other societies belonged to associations which fall into two or more subject matter categories, and 12 per cent were members of societies classified in three or more of the general groups. Three hundred members (16 per cent) held col-

lateral memberships in an economic, political, or other social science association; almost 9 per cent were associated with both a political and social science association and an organization classified under "other associations"; and 7 per cent were members of both "economics" and "business management" groups.

Table 8 shows collateral memberships held by American Statistical Association members in any two of fourteen other professional societies which were checked most often by reporting members. For example,

TABLE 8

OVERLAPPING MEMBERSHIPS OF AMERICAN STATISTICAL ASSOCIATION MEMBERS' AFFILIATION WITH ONE OR MORE OF 14 OTHER PROFESSIONAL SOCIETIES IN 1945

	Professional Societies	Number and Percent of A.S.A. Members Affiliated with Both Societies														
		15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
		A.S.A.	S.A.M.	P.S.	A.F.E.A.	A.P.A.	A.S.S.	P.A.A.	A.P.H.A.	A.P.S.	E.S.	A.A.P.S.S.	A.M.A.	A.A.A.S.	I.M.S.	A.E.A.
1	A.E.A.	729	19	3	40	3	10	30	12	103	106	101	96	92	44	
2	I.M.S.	313	3	20	2	13	9	11	14	6	56	13	10	60		2.4
3	A.A.A.S.	272	1	28	6	33	32	41	44	24	30	30	16		4.3	3.3
4	A.M.A.	216	16	4	7	5	5	3	3	22	10	23		0.9	0.5	5.2
5	A.A.P.S.S.	196	4	9	6	1	23	20	22	46	17		1.2	2.1	0.7	5.4
6	E.S.	173	2	12	10	6	5	12	4	21		0.9	0.9	2.1	3.0	5.7
7	A.P.S.	152	3	1	7	2	4	7	6		1.1	2.5	1.2	1.3	0.3	5.6
8	A.P.H.A.	140	0	1	0	4	13	33		0.3	0.2	1.2	0.2	2.4	0.8	0.6
9	P.A.A.	134	2	4	3	3	56		1.8	0.4	0.6	1.6	0.2	2.2	0.6	1.6
10	A.S.S.	102	3	4	3	5		3.0	0.7	0.2	0.3	1.2	0.3	1.7	0.5	1.0
11	A.P.A.	77	1	25	0		0.3	0.2	0.2	0.1	0.3	0.1	0.3	1.8	0.7	0.2
12	A.F.E.A.	76	0	0		—	0.2	0.2	—	0.4	0.5	0.3	0.4	0.3	0.1	2.2
13	P.S.	57	1		—	1.3	0.2	0.2	0.1	0.1	0.6	—	0.2	1.5	1.0	0.2
14	S.A.M.	50		0.1	—	0.1	0.2	0.1	—	0.2	0.1	0.2	0.0	0.1	0.2	1.0
15	A.S.A.	1854	2.7	3.1	4.1	4.2	5.5	7.2	8.0	8.2	9.3	10.6	11.6	14.7	16.9	39.3

NOTE: Percentages are shown in the lower half of this table and can most easily be related to the number of members by use of row and column numbers. The complete name of each professional society can be determined by comparing its code number with the code numbers given in Table 7.

of the 729 members who belonged to the American Economic Association, 106 also belonged to the Econometric Society and 103 were members of the Academy of Political Science. Duplications exist in these figures.[6] Among those societies not shown on this table, the most important overlapping occurred between the two actuarial societies listed, 25 American Statistical Association members belonging to both organizations.

[6] It is not possible to present in this paper the many combinations of societies to which American Statistical Association members belonged. However, data for any particular combination are available and can be had by writing to the National Office.

EDUCATION

More than 92 per cent of the American Statistical Association members have received degrees from colleges or universities, the highest degrees attained being divided almost equally between bachelor of arts, master of arts, and doctor of philosophy or their equivalents.

Fifty-seven per cent of the membership had degrees in one of the

TABLE 9

EDUCATIONAL MAJORS AND DEGREES OF AMERICAN STATISTICAL ASSOCIATION MEMBERS IN 1945

Education Major	Members Who Reported That They Received Degrees		Highest Degree Received or Equivalent					
			B.A.		M.A.		Ph.D.	
	Number	Per Cent	Number	Per Cent	Number	Per Cent	Number	Per Cent
All A.S.A. members reporting degrees*	2563	100.0	861	33.6	910	35.5	792	30.9
Economics	915	35.7	266	10.4	307	12.0	342	13.3
Mathematics	339	13.2	160	6.2	100	3.9	79	3.1
Statistics	269	10.5	92	3.6	122	4.8	55	2.1
Business administration and marketing	161	6.3	55	2.1	96	3.7	10	.4
Sociology	110	4.3	19	.7	39	1.5	52	2.0
Engineering	106	4.1	78	3.0	23	.9	5	.2
Psychology	90	3.5	11	.4	22	.8	57	2.2
Social science	69	2.7	23	.9	27	1.1	19	.7
Liberal arts	57	2.2	44	1.7	12	.5	1	†
Education	54	2.1	10	.4	24	.9	20	.8
Chemistry	54	2.1	19	.7	10	.4	25	1.0
Biology	46	1.8	14	.5	7	.3	25	1.0
Medicine	42	1.6	1	†	—	—	41	1.6
Political science	41	1.6	9	.4	17	.7	15	.6
Law	37	1.4	—	—	37	1.4	—	—
Agriculture	31	1.2	8	.3	13	.5	10	.4
Public health	26	1.0	2	.1	11	.4	13	.5
Physics and geology	22	.9	7	.3	6	.2	9	.4
Other studies	94	3.7	43	1.7	37	1.4	14	.5

* 208 of the 2724 members who reported education did not receive college or university degrees.
† Less than .05 per cent.

social sciences or their specialized disciplines; 65 per cent of those who held doctorates had a major in the social sciences. As would be suggested from previous tables, most of the members not only have B.A. degrees in economics, but one in eight have doctorates in this field.

One-fourth of the members have educational majors of mathematics and statistics; one in seven has doctorates in these two fields. Table 9 shows the major field of study for those members who reported that they had received one or more degrees.

PARTICIPATION OF THE MEMBERS IN THE ACTIVITIES OF THE AMERICAN STATISTICAL ASSOCIATION

One out of every seven members has had papers or articles published in the JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION. However, more than half have published articles in other journals or books. Thirteen per cent of the membership have given papers at annual meetings and are in general the same members who wrote for the JOURNAL. Four per cent of the membership have held one or more national offices, and seven per cent have been officers in one of the local American Statistical Association chapters. Three per cent indicated that they had been chairman of one or several sessions of annual meetings.

GEOGRAPHICAL DISTRIBUTION OF THE MEMBERSHIP OF THE AMERICAN STATISTICAL ASSOCIATION

The change that has occurred in the geographical distribution of the Association's membership over the years is shown in Table 10. The per

TABLE 10

SHIFT IN GEOGRAPHICAL DISTRIBUTION OF AMERICAN STATISTICAL ASSOCIATION SINCE 1893

Area	1893	1920	1940	1945
Number of members	504	920	2044	3307
Per cent of members in				
New England	36.4	11.9	9.4	4.9
Middle Atlantic	28.4	45.8	35.6	31.5
South Atlantic	10.5	13.4	20.8	31.2
East North Central	12.7	18.6	18.8	12.4
Other areas	12.0	10.3	15.4	20.0

Data for 1893, 1920, and 1940 from table in article by Funkhouser, *op. cit.*, p. 330.

New England: Me., Vt., N. H., Mass., R. I., Conn.; Middle Atlantic: N. Y., Pa., N. J.; South Atlantic; W. Va., Va., Md., Del., N. C., S. C., Ga., Fla., D. C.; East North Central: Wis., Ill., Ind., Ohio, Mich.

cent of the membership in the New England region has declined steadily since 1893, while the proportion of the members in the South Atlantic region, which includes the metropolitan area of Washington, has continued to increase. In 1945 about 63 per cent of the members were located in the Middle and South Atlantic regions; 27 per cent lived in the Washington, D. C. area; and about one-fourth were located in New York state, chiefly in the City of New York. Women, comprising ten per cent of the total membership were distributed fairly proportionately among the areas.

The tables presented in this paper summarize the more important characteristics of the American Statistical Association membership. In the process of compiling these summaries all available information was coded and placed on punch cards, thus making further summarization or detailed analysis possible without undue additional effort. The file of punch cards will be maintained for six months and every effort will be made during that period to satisfy any considered request of individual members for additional information.

BRITAIN'S NEW POST-WAR ECONOMIC GUIDE

Norman Crump
Financial Editor "Sunday Times" (London)

Nearly two years ago Britain's Government of the day, which included representatives of all political parties, issued a White Paper on Employment Policy. The concluding passages of this document made it clear that all the Government's proposals for the maintenance of employment depended for their success upon a full and continuous knowledge and appreciation of all facts bearing upon the nation's economic position. It, therefore, went on to say that "the Government intend to establish on a permanent basis a small central staff qualified to measure and analyze economic trends and submit appreciations of them to the Ministers concerned."

Actually such a staff came into existence early in World War II. It is described as the Central Statistical Office, and was directly responsible to the War Cabinet. Much of its work necessarily remained secret during the war, but it was responsible for the annual White Paper measuring and analyzing Britain's national income, as well as for the document describing Britain's war effort, which appeared towards the end of 1944.

Quite rightly, it is to remain a permanent body. One of the first fruits of its post-war labors is a new publication called the *Monthly Digest of Statistics*. This is available to the public, as well as to Ministers and Government offices. The first issue, for January, 1946, appeared in the middle of February.

Part of the information which it contains was already appearing elsewhere. For years past the Bank of England had published a monthly *Statistical Summary*, which was of the utmost use to those concerned in economic trends. Publication of this *Summary* has now ceased, and many of the statistics which it included have been transferred to and will be continued by the new publication. Similarly the *Board of Trade Journal* publishes statistics of wholesale prices, external trade, coal and steel production, etc. These again are to be found in the new document. So, too, with the statistics of wages, earnings, employment and the cost of living which appear each month in the Ministry of Labor's Gazette.

Here the main advantage is that these statistics are assembled in a single book. How great an advantage that will prove will be realized by all those who have to look up and interpret a particular statistical fact, while working against the clock. But the Digest does much more than assemble material which either is or was available elsewhere. It

includes many new tables, which were originally compiled for Government use during the war, but which for security reasons have hitherto remained unpublished.

The *Digest* falls into eleven main sections. These are Employment; Fuel and Power; Raw Materials (production, consumption and stocks); Manufactured Goods; Building; Agriculture and Food; External Trade; Merchant Shipping; Inland Transport; Finance (Government and Banking); and Wages and Prices. The tables are not yet entirely complete. In some instances stocks are not included. In others, pre-war statistics are either lacking or have been omitted. It is to be hoped that this will be remedied in subsequent issues, but in some cases it must be recognized that the necessary details only began to be collected during the war.

These are small points of criticism. It is much more important to notice that the raw materials section includes iron and steel in considerable detail; non-ferrous metals; timber; textiles (production, consumption and stocks); hides, skins and leather; rubber, both natural and synthetic; chemicals; and wood-pulp and paper. Manufactured goods statistics range from motor cars to sewing machines and electric irons; from fountain pens to tooth brushes; from dustbins to footwear; and wardrobes to dining-room chairs. There are details of beer, spirits and tobacco, wireless licences and car registrations; retail sales of food, clothing and household goods, in index form. The Building section includes the number of operatives at work, production of materials and components, temporary houses, and the progress (still in the initial stages) made towards the building of permanent houses.

Of greatest importance perhaps is the section on Employment. This is good as far as it goes. But the latest statistics are those for November, 1945, and demobilization has since made further considerable progress. Nor does the *Digest* contain any attempt to look ahead. There is nothing resembling the "man-power budget," advocated in the Employment Policy White Paper as an instrument essential to the proper formulation of employment policy.

That is not the fault of the Central Statistical Office. The preparation of a man-power budget depends upon decisions of the highest policy. But until those decisions are taken, and the man-power budget is prepared and published, the new *Statistical Digest* can only be an auxiliary, even though invaluable within its own limits. It will not by itself enable either Government Departments or industrialists to complete their future plans.

PROBLEMS AND METHODS OF THE SAMPLE SURVEY OF BUSINESS*

MORRIS H. HANSEN, WILLIAM N. HURWITZ AND MARGARET GURNEY
Bureau of the Census

In business sampling, as in the general sampling problem, the over-all test that we apply to a sample design is that it shall yield the desired information with the reliability required at a minimum cost; or, conversely, that at a given cost it shall yield the estimates desired with the maximum reliability possible. A second criterion which we impose on the sample design is that the reliability of the sample results should be measurable. This requirement eliminates from consideration many superficially attractive sampling schemes which, while generally economical in both time and money, may lead to faulty conclusions.

A particularly important characteristic of the distribution of sales of business establishments is that the distribution is highly skewed, with relatively few large establishments accounting for a substantial proportion of total sales. The sample design takes advantage of this fact and in so doing yields a sample of greater efficiency than would result from a sample of equal size if the distribution were normal. Various size classes are distinguished and the theory of optimum allocation indicates the number of stores to be sampled from each class. Large stores will be sampled from a listing of establishments, and the more numerous smaller stores will be sampled by cluster sampling.

The appropriate sampling theory is given for measuring the various contributions to the sampling error, and for maximizing the information obtained per dollar expended.

THE APPLICATION of recent advances in sampling methods to the problem of measuring, for example, retail sales and changes in retail sales by kinds of business now makes it possible to improve the quality of the current survey work being done in the business field by the Bureau of the Census. The revision of the survey methods will be carried out in connection with the expansion of this work now planned. The revision and expansion of the survey involves designing a sample of retail stores that will produce: (a) reliable national estimates of total retail sales and of month to month changes in sales, by kind of business; and (b) similar figures with somewhat less precision for individual large cities, and subsequently, for regions and individual States. In this

* This paper was presented at the annual meeting of the American Statistical Association on January 26, 1946 in Cleveland, Ohio. The authors wish to thank Mr. Max A. Bershad for reviewing the manuscript and making numerous helpful suggestions.

paper we shall consider primarily the design of the sample to meet the first of these objectives. The extension of the sample to provide statistics for individual cities and for other areas will be made in such a way as to further improve the general sample, but these methods will not be discussed here.

SOME FUNDAMENTAL CONSIDERATIONS IN SAMPLE DESIGN

In business sampling, as in the general sampling problem, the overall test that we apply to a sample design is that it shall yield the desired information with the reliability required at minimum cost; or conversely, that at a given cost it shall yield the estimates desired with the maximum reliability possible. Various resources may be available, and restrictions and limitations other than cost restrictions may necessarily be imposed on the design. The resources may be in the form of available statistical data, lists of establishments, an existing field organization, tabulation facilities, existing knowledge of the problem being dealt with, etc., together with available mathematical statistical tools. The restrictions and limitations may have to do with time limits to be met, limitations on kinds of personnel that can be used, etc. Designing the sample to achieve the maximum information at minimum cost involves, of course, making fullest possible use of resources and operating within the necessary restrictions.

A second criterion that we impose on sample designs in the Bureau is that the reliability of the sample results should be measurable. Methods of sample selection and estimation are available for which the risk of errors in the sample estimates can be measured and controlled. As the size of the sample is increased, the expected discrepancies between the estimated value from the sample and the true value (i.e., the value that would be obtained from a complete census) will decrease if such methods are used. With such methods one can know the risk taken that the error due to sampling will exceed a specified amount. This risk (i.e., the risk that the error will exceed any specified amount) can be made as small as desired by taking a sample of adequate size.

The use of methods yielding results whose sampling errors can be measured has these advantages: we can indicate to the users of our data the degree of confidence they can place in the published figures; we can make use of an objective criterion for choosing the "best" from among alternative designs; and we can determine in advance of drawing the sample whether resources are available to get results from the sample with the reliability required, and so we may be able to prevent useless sample surveys from being carried forward.

If we permit the use of sample designs that do not provide measurable results, a great many superficially attractive sampling schemes are eligible for selection. For obtaining a sample of a given size these methods are generally economical, in both time and money, but they may lead to faulty conclusions.

In referring to sampling errors or reliability of sample results here the reference, of course, is to the discrepancy between a sample result and the results that would have been obtained from a complete census carried out with equal care. Problems of accuracy of response are thus not being considered in this paper. It should be pointed out, however, that in practice it may be possible with a sample to carry out more careful operations and insist on greater accuracy in the various field steps than is feasible in a complete census, so that actually nonsampling errors may be reduced in a carefully conducted sample survey below what would be present in a complete census.

DIFFICULTIES OF METHODS OFTEN USED IN BUSINESS SAMPLING

A method often used in business sampling is to mail questionnaires to all or a sample of establishments included on available lists and to depend primarily on a voluntary mail response to the mailed inquiry. In using such returns it has usually been assumed that, if sales for two periods of time are collected from an identical set of stores, the change in sales between the two periods for the identical stores will, in fact, reflect the change in total sales for the specified kind of business. This type of sample estimate may be subject to two sources of bias: (1) the mail responses may not reflect the change in sales of the nonrespondents to the mail questionnaire; (2) the available mailing lists are usually out of date and hence the sample drawn from the lists does not provide the information for estimating the net effect of the turnover in business establishments.

It is possible to overcome the kind of difficulties outlined by the use of modern sampling methods. Using these methods we can produce results the reliability of which can be controlled, and which, with an adequate size of sample, can be made as precise as desired. At the same time, these methods point the way to minimizing the cost of obtaining results with the desired precision.

CHARACTERISTICS OF THE POPULATION TO BE SAMPLED AND RESOURCES AND FACILITIES AVAILABLE

A particularly important characteristic of the distribution of sales of business establishments is that the distribution is highly skewed. A

relatively few large establishments, constituting a small proportion of all establishments, account for a substantial proportion of the total sales. At the same time, the very large number of remaining establishments also account, in the aggregate, for a substantial proportion of total sales. This highly skewed distribution can be utilized in the sample design to produce more reliable results for a given size of sample than would be possible, if, say, the sales were normally distributed. The advantage of the skewness of the population can, of course, only be utilized if approximate measures of size are available in advance of taking the survey, thus making it possible to select the larger establishments for special treatment in sampling. This type of information is available for the larger establishments, although no up to date and complete list of all retail establishments is available.

Records that are available and will be used include: (a) Social Security records—the Social Security Board has records of employers in covered industries having one or more employees. These records are available on a relatively up to date basis and are classified by primary kind of business. (b) The last census of business—the 1940 Census provides a record of establishments by kinds of business as of the date of the census, although these records are not now up to date. However, the census provides extensive statistical information on retail trade as of 1940 that can be effectively utilized.

In addition to the Social Security records and the data from the 1940 Census, another important resource available to the Bureau of the Census is an existing field organization operating in 68 different sample areas. This field organization has been established for purposes of taking the Monthly Report on the Labor Force and can be used and extended for the current business survey.

An additional resource of importance is the availability in the Census of comparatively up to date maps covering cities having more than 25,000 population. These maps are brought up to date annually and show, for the most part, the location and general type of use of individual structures. They have been acquired by the Census as an aid in sampling and in census taking, and can be effectively utilized in sample surveys of both population and business.

We shall not attempt to present a detailed description of the sample, as many of the details will be modified as the design is put into effect, but a brief description of the sampling plan will be given with elaboration of one or two specific points as illustrative of the role that modern sampling methods play in determining the sample design.

DIVISION OF RETAIL TRADE INTO TWO SUBUNIVERSES AND ALLOCATION OF SAMPLING TO EACH

It will be convenient to regard the universe of all retail stores in the United States as divided into two subuniverses: first, those that are included on a list of large stores that will be prepared; and second, those that are not. One can proceed separately to design a sample for each of these subuniverses.

A question that must be considered concerning the subdivision of the population into a subuniverse of listed large stores and a subuniverse of all other stores is: How many establishments should be included on the large store list? Records are available from which more than half of all stores could be included on the list, whereas it is expected to include only about the largest 10 per cent. Why include only 10 per cent? Certain optimum sampling principles, which will be described later, dictate that the very large stores must be on the list and suggest that the more stores included on the list the better. On the other hand, various administrative considerations, such as the difficulty of identifying and matching stores in order to avoid including some in both subuniverses, together with problems of dealing with births and deaths of establishments, operate to hold the number of stores included on the list to a minimum. At the outset, these administrative factors point to keeping the list down to the 10 per cent now contemplated.

The list of large establishments, classified by kind of business, has been prepared based on the Social Security records. Each establishment on the list has an indication of size based on payrolls during a previous period. Of course, the actual current size of each establishment listed is not known, but the wages paid during a prior period provide an approximate measure of size.

The use of past payroll information to distinguish large from small stores means that there will be a number of stores not on the list that will be larger in current sales than many of those that are on the list. Despite this fact the use of a subuniverse of listed large stores will be highly effective in increasing the reliability of the sample. With the methods to be followed, if the measures of size used are highly correlated with the actual sales during the current periods under consideration (as will be the case), then the use of the measures of size in classifying the establishments will be exceedingly effective in reducing the sampling error. On the other hand, even if the measures of size were not highly correlated with actual sales, no bias would be introduced into the sample. Thus, the design is such that effective use has been

made of available resources so that their use increases the sample efficiency without in any way biasing the sample.

Considerations of optimum allocation[1] make it clear that the subuniverse of listed large establishments should be sampled more intensively than those not on the list. Consequently, approximately 30 per cent of those on the list will be included in the sample, while only about 2 per cent of those not on the list will be included. This allocation of sampling between the two subuniverses arises because the stores listed, while constituting only about 10 per cent of the total stores, account for about 60 per cent of the total sales. Conversely, those not on the list account for about 90 per cent of the stores and 40 per cent of the total sales. Actually, by including about 30 per cent from those on the list, and about 2 per cent from the remaining stores, we find that the number of establishments included in the sample from each of the classes will be roughly proportionate to the aggregate sales of each, rather than proportionate to the number of stores. This is in accord with optimum sampling principles where the principal characteristics to be estimated from the sample are total sales or changes in total sales.

To illustrate this principle, suppose we have stratified the population of establishments into two strata, and want to allocate the sampling to each stratum in such a way as to produce the most reliable results possible per dollar of cost. Suppose, also, that we want to estimate the ratio of sales in the current month to those in the prior month from the sample. We shall use as our estimate the ratio

$$\frac{\dfrac{N_1}{n_1} X_1 + \dfrac{N_2}{n_2} X_2}{\dfrac{N_1}{n_1} Y_1 + \dfrac{N_2}{n_2} Y_2}, \tag{1}$$

where X_1 represents, for the first stratum, the aggregate sales during the current month of the stores in the sample, and X_2 is the corresponding figure for the second stratum. Similarly, Y_1 and Y_2 represent, for the two strata, respectively, the aggregate sales of the same stores during the prior month. The total number of stores in the first stratum is N_1, and n_1 is the number of these included in the sample, while N_2 and n_2 are the corresponding figures for the second stratum. The vari-

[1] The notion of optimum allocation of a sample was introduced by J. Neyman, see "On the Two Different Aspects of the Representative Method; A Method of Stratified Sampling and the Method of Purposive Selection," *Jour. Royal Stat. Soc.*, New Series, Vol. 97 (1934), pp. 558-606. See also P. C. Mahalanobis, "A Sample Survey of the Acreage under Jute in Bengal," *Sankhyā*, Vol. 4 (1940), pp. 511-530.

ance of this sample estimate is given approximately by:

$$(2) \qquad \sigma^2_{\text{ratio}} = \frac{N_1 - n_1}{(N_1 - 1)n_1} \frac{Y_1^2}{Y^2} S_1^2 + \frac{N_2 - n_2}{(N_2 - 1)n_2} \frac{Y_2^2}{Y^2} S_2^2.$$

The first term above represents the contribution of the first stratum to the over-all variance. In this term, Y_1 is the aggregate sales of all establishments in the stratum during the prior month and S_1^2 is roughly equivalent to the variance between individual establishments in that stratum of the "sales ratio." By the "sales ratio" is meant the ratio, for an establishment, of the sales in the second period to those in the first. The second term has a similar meaning for the second stratum. Finally, $Y = Y_1 + Y_2$; that is, Y is equal to the aggregate sales during the prior period of all establishments in both strata combined.[2]

We want to obtain a sample that produces results of specified reliability at minimum cost, and to do this we must take account, also, of the relative cost of including an establishment in the sample from the first stratum as compared with the cost of including one from the second. Suppose, therefore, that the cost of including an establishment from the first stratum is k times the cost of including one from the second (k may be either greater or less than 1).

With these facts we can now ascertain the relative size of sample that should be taken from each stratum so as to produce results of the maximum accuracy per dollar of cost. We accomplish this by setting

$$(3) \qquad \frac{n_1}{n_1 + n_2} = \frac{Y_1 S_1}{Y_1 S_1 + Y_2 S_2 \sqrt{k}}.$$

It is apparent from this that if k is approximately equal to 1 (that is, if the cost of including an establishment in the sample is about the same

[2] Actually, $S_1^2 = \sigma_1^2 + (R_1 - R)^2 V_{1Y}^2 + B_1$, where $\sigma_1^2 = (V_{1X}^2 + V_{1Y}^2 - 2\rho_1 V_{1X} V_{1Y}) R_1^2$ is approximately the variance of the sales ratio for establishments in the first stratum; $R_1 = \frac{X_1}{Y_1}$, and $R = \frac{X_1 + X_2}{Y_1 + Y_2}$, with X_1 equal to the aggregate sales of all establishments in the first stratum during the current month, and X_2 similarly defined for the second stratum;

ρ_1 is the correlation between sales in the two periods for establishments included in the first stratum;

V_{1X}^2 is the coefficient of variation of the sales during the current period of the establishments in the first stratum, and

V_{1Y}^2 is the corresponding coefficient of variation for the prior period;

$B_1 = 2(R_1 - R)R_1(\rho_1 V_{1X} V_{1Y} - V_{1Y}^2)$, and B_1 will usually be of negligible size relative to the other terms of S_1^2.

S_2^2 is similarly defined but relates to establishments in the second stratum.

If more than two strata were used the variance of the sample estimate would be for, say, L strata

$$\sigma^2_{\text{ratio}} = \sum_{i=1}^{L} \frac{N_i - n_i}{(N_i - 1)n_i} \frac{Y_i^2 S_i^2}{Y^2},$$

where the terms for the ith stratum would all be defined as above for the first stratum.

for both of the strata), and if $S_1^2 = S_2^2$, which is approximately true when the standard deviation of the sales ratio is the same for establishments in the two strata, the optimum proportion of stores to be included in the sample from each stratum is equal to the proportion of the aggregate sales accounted for by the establishments in the respective strata.

It is such considerations that determined the relative allocation of sampling indicated earlier between the subuniverse of listed large establishments and the remaining stores. Moreover, it will be found that by using this optimum allocation method, the sampling error will be reduced very substantially below what would result from proportionate sampling from each subuniverse. Of course, not all of the facts necessary to determine an optimum job are available in advance, and adjustments will be made in the sample with further experience. However, knowledge of the principles and of the general characteristics of the population to be sampled makes it possible to approximate the optimum in the initial design.[3]

The theory just presented not only indicates the optimum sampling ratio to be applied to each of two size strata, but suggests that still more gains might be achieved by additional stratification by size of establishment and by using increasingly higher sampling ratios as the size of store increases. Such considerations suggest, also, that all stores larger than some particular size are of sufficient importance that they should be included in the sample. This theory, therefore, has a bearing on the question discussed earlier concerning how many stores should be included in the subuniverse of listed large stores. The relevance of the theory just presented to this latter issue arises because, while it is feasible to use sampling ratios ranging from 100 per cent down in sampling from various classes of stores included on the large store list, the method of sampling from the remaining stores is such that only small ratios are practicable. In fact, it is anticipated that at the outset not more than a 2 per cent sampling ratio will be applied to any sizes or classes of stores not included on the large store list.

THE SAMPLE FROM THE SUBUNIVERSE OF LISTED LARGE STORES

In the previous section we described how retail stores were divided into two subuniverses, a list of large stores, and all stores not on the

[3] It should be recognized that the theory of optimum sampling just outlined holds for any single statistic, but where estimates of a number of different items are wanted from a sample a compromise may have to be made between the designs which would be optimum for each of the various statistics involved. Often, however, as is the case here, many of the items for which estimates are wanted will be highly correlated, and for these the design is close to the optimum for each.

list. We shall now consider the sample from the list of large stores. The variation among the sizes of even those stores on this list is very great and consequently one should not take a uniform sampling ratio from these if he is to approximate an optimum sample design.

The principles discussed in the previous section suggest that optimum allocation would call for including in the sample all stores having sales larger than a certain size, and that the large store list should be divided into a number of strata based on size of store with a different sampling ratio assigned to each. There are, however, serious disadvantages in having too many different sampling ratios. The more sampling ratios imposed, the larger the number of different weights that must be used. Moreover, most of the gains of optimum allocation to size groups will be achieved if we set up only two classes of large stores, one consisting of the stores which will be included with certainty, and the other consisting of the remaining listed large stores. Once the list of large stores is prepared, therefore, it will be further divided into a *very large store list* from which we make every effort to obtain returns 100 per cent, and a *medium large store list* from which perhaps 20 per cent returns may be required.

Joint use of mail and personal follow up. Since each of the very large stores is of sufficient size that it should be included in the sample, no matter where located, this small group will be mailed to and followed up with field interviews as necessary to insure substantially complete coverage each month. The follow-up work will be facilitated by the fact that most of the very large establishments are concentrated in the larger cities and that a prompt mail response can be expected from most stores in this group.

A further application of the principle of optimum use of methods is made in the joint use of mail canvass and field follow-up methods in sampling from the medium large store list. The collection of a questionnaire by mail can be done at a fraction of the cost of collection by field interview—provided a mail response to the questionnaire can be obtained. Consequently, the fullest possible use of this method of canvass should be utilized with minimum use made of the much more expensive method of personal follow up.[4] The schedules to be mailed each month will be in excess of the number actually expected to be returned, and enumerative follow-up will be made of a sample of those that do not respond to the mail canvass. The plan is for a rotation

[4] See "The Problem of Non Response in Sample Surveys" (this JOURNAL) for a description of optimum methods of joint use of mail and enumerative response, or of any two methods where one is less expensive but the other is more certain of obtaining response.

in the follow-up of the medium large stores not responding so that over a period of a relatively few months all of those not responding to the mail canvass will be visited in an effort to build up cooperation to the mail survey. With the mail response and coverage through follow-up of a sample of those not responding, it is possible to compute sample estimates without assuming that those responding are representative of all medium large establishments.

Use of cluster sampling to facilitate field work by restricting sample to selected primary areas. Field follow-up of nonrespondents to a mail survey would be exceedingly expensive if the establishments to be covered were widely scattered throughout the entire United States. If it were not for this expense the optimum procedure would call for designating the establishments to be included in the sample with no clustering of the sample establishments whatever. To deal with the problem of cost, however, the device of clustering the sample will be used. A sample of primary areas will be designated and mailing to the medium-large stores will be limited to those located within the selected primary areas. A cross-section of such areas will be designated throughout the United States such that, if the retail stores within these areas are adequately represented in the sample, sufficiently reliable information for all stores will be available. This does not mean that the sample should be adequate to represent well each primary area separately. It does mean that the sample should be adequate to represent all primary areas combined.[5] Since the list of medium-large stores is not extensive, all the establishments on this list within these areas will be mailed to, and the follow-up of a sample of those not responding can be instituted. Those to be contacted will be rotated so that over a period of perhaps three months all nonrespondents to the mailed questionnaires will be visited.

Optimum choice of primary areas. With the adoption of this cluster sampling approach the question still remains: What kind of a sample of areas ought we to select? The solution to this problem was fairly clear in this instance because, as indicated earlier, the Bureau has an operating field organization taking the Monthly Report on the Labor Force. A somewhat similar problem presented itself in the design of the labor force sample. For that survey, a sample of 68 primary areas has been designated within which subsampling operations are carried on, and a full-time supervisor with part-time enumerators is available in each of these areas. Each primary area consists of a county, or, in some

[5] This important principle is often misunderstood. See, for example, the paper entitled "Design, Size, and Validation of Sample for Market Research," *Journal of Marketing*, January 1946, by the Committee on Marketing Research Techniques.

instances where the sizes of the counties are not too large, of two or three adjacent counties. The principles of sampling and the procedures followed in designing the labor force sample, described elsewhere,[6] are equally applicable to business sampling.

An obvious question to be answered before the adoption of these sample areas for business sampling is: What reason is there to presume that a sample of areas designated for such a population sample will prove to be a satisfactory set of areas in which to take a retail trade sample. The answer is not too difficult. In the first place, the design of the population sample was an unbiased design, from which it follows that unbiased estimates of both population and business as well as of other characteristics can be obtained by a survey of these areas. The real question then is: Will this sample produce results having a sufficiently small sampling error? Since retail trade exists to serve the population, it is not surprising to find that it is distributed between counties much as the population is distributed and, therefore, that a sample designed to estimate the labor force characteristics will be comparatively efficient also for estimating retail trade, if appropriate methods of estimation are used. It can be seen from the following table that the average sampling errors resulting from the use of a sample designed exactly as the labor force sample but used for retail trade will produce national estimates for many kinds of business with fairly high precision.

NATIONAL ESTIMATES OF RETAIL SALES FOR ALL RETAIL TRADE AND FOR A FEW KINDS OF BUSINESS FROM THE 68 AREA LABOR FORCE SAMPLE*

	Sales			
	Known 1939 Census	Estimate from sample	Per cent difference	Coefficient of variation of sample estimate in per cent
Total—all types	42,041,700	42,170,007	+0.3	1.4
General Stores (with food)	810,342	791,005	−2.4	12.6
General Mdse. Group	5,665,007	5,678,947	+0.2	2.5
Drugs	1,562,562	1,533,235	−1.9	2.0
Apparel	3,258,772	3,187,103	−2.2	2.0
Lumber-Hardware	2,734,914	2,563,700	−6.3	3.9
Furniture	1,733,257	1,657,040	−4.3	3.3

* Computations for this table were made for a special survey which covered only these six kinds of business. General Stores (with food) occur in the rural areas, but seldom in larger cities, and consequently, for the 68 area sample the sampling error for this special class is large.

[6] Morris H. Hansen and William N. Hurwitz, "A New Sample of the Population," *Bureau of the Census*, September, 1944.

This table gives an estimate of the sampling errors to be expected in estimating total sales for a few kinds of business from the 68 labor force sample areas—but on the assumption of complete coverage of stores within the designated areas. Thus, the table provides a guide to how representative these areas are of the entire nation with respect to retail trade. The per cent difference column shows the actual errors in estimating total sales from the sample counties, while the coefficient of variation indicates the average error to be expected over all possible samples drawn by the method actually followed.

We see that the labor force and general population sample may be expected to produce retail trade estimates of reasonable reliability, but it should be clear that at least some gains in precision, whether small or large, could be achieved if a sample of areas were utilized that was designed explicitly for retail trade rather than for some other problem. The question arises, then, why not design a new sample? The answer to this question has two aspects. The first is that, for the same size sample, some gains undoubtedly could be achieved by a more specialized design. Nevertheless, we have had considerable experience indicating that, on the average for the many different statistics to be estimated, the gains which would result would not be large. The second and principal point is that, with field facilities already existing in these counties, the cost of obtaining returns is substantially less by utilizing the available counties than by drawing an independent sample in other areas. Therefore, *per dollar*, the most reliable results are obtained in the way specified. If more reliable results are wanted, we need merely to expand that sample. Such considerations led to the adoption of the 68 area labor force sample as the areas in which mailing and field operations would be carried on; this sample subsequently will be enlarged by expanding to additional areas.

Principles involved in the determination of the number of primary areas to be covered and the size of sample in each. The proposed expansion to more areas suggests the need of some guide as to how many primary sampling areas should be listed in the sample and how large a sample should be selected within each. We shall indicate some principles that serve as a guide in making this determination, but shall present them in a very simplified problem. In particular, we shall consider only the problem of sampling from a single stratum, and we shall assume that the number of stores is the same for all counties in the stratum. The principles pointed up and conclusions arrived at for this simple situation are equally applicable to the more complex problem with which we are actually dealing.

Suppose that there are a total of M counties in a stratum, from which we take a sample of m at random; that there are a total of N stores in each county in the stratum, and that we subsample n from each selected county. Thus, we will have a sample of m counties from this stratum, and of n stores from each county, or a total sample of mn stores. If we compute the average sales per establishment for the stores in the sample the variance of this sample average will be:

$$\sigma^2_{\text{average}} = \frac{M-m}{(M-1)m}\left\{\sigma_b^2 - \frac{M-1}{M}\,\frac{\sigma_w^2}{N-1}\right\} + \frac{MN-mn}{M(N-1)}\,\frac{\sigma_w^2}{mn}$$

where σ_b^2 is the variance of the average sales per establishment between counties in the stratum, and σ_w^2 is the average within county variance.

The term $\left\{\sigma_b^2 - \dfrac{(M-1)\sigma_w^2}{M(N-1)}\right\}$ will, with very few exceptions, be positive in practical sampling problems. For a fixed total size of sample, that is, for fixed mn, the second term of the expression for the variance of the sample average will be unaffected by the way in which the sampling is divided between m and n. Therefore, for a sample of the total size of mn, the variance of the sample average will be smallest when the maximum possible number of counties is included in the sample and the sampling within counties is correspondingly reduced. Thus, if a sample of 500 stores is to be taken, i.e., if $mn = 500$, the most reliable result will be achieved by taking only a single store from each of 500 sample counties. It follows that if there are not enough counties in the stratum to accomplish this, then the optimum sampling would call for including *every* county in the sample.[6]

With this fact established, why then do we propose to take a relatively small number of counties into the sample with a considerable number of stores from each? The reason is that we are trying to get the most reliable result possible per dollar spent, rather than per store included in the sample, and considerable economies can be effected by clustering the sample.

To illustrate this, let us go one step farther and assume that the total cost, C, of taking a sample can be broken down as follows:

$$C = c_1 m + c_2 mn.$$

This amounts to saying that there is an overhead of C_1 dollars that is incurred for each county that is included in the sample, and that a cost of C_2 is incurred for each store covered within a sample county.

This, again, is a much oversimplified assumption, but it will be useful in showing how the allocation of sampling as between including more counties and taking a larger sample within fewer counties may be affected by cost considerations.

Now, if we find the size of sample to be taken in each county that will produce the most reliable results possible per dollar spent we get

$$n = \sqrt{\frac{C_1}{C_2}} \frac{\sigma_w}{\sqrt{\left(\sigma_b^2 - \frac{\sigma_w^2}{N-1}\right)}} \tag{5}$$

as the sample to take within each county. Then m, the number of counties to be included, is the number that the budget allows or that is needed for the required accuracy. This theory provides a guide for determining how many counties to sample and how large a sample to include from each. As we learn more about the costs involved and the population we are sampling and its distribution between counties, this theory will be available to improve the sample design. At the outset, we know enough based on research already done and experience and general knowledge of retail trade to determine the first approximation that should be followed. That first approximation, in this case, is to start out initially with the sample of 68 primary areas used for the labor force sample, and to take a fairly large subsample within each primary area, although it is to be expected that the sample will be expanded to more areas subsequently.

THE SAMPLE FROM THE SUBUNIVERSE OF SMALLER STORES

We have up to this point a sample from only the subuniverse of listed large stores. For stores not on the list, which include the great bulk of establishments, we propose to confine our sample to the same primary areas that are being used for sampling the listed medium-large stores. In addition to the advantages of minimizing cost of travel already discussed, the use of these primary areas for the remaining stores makes it feasible to develop and maintain an up to date sample of births and deaths of stores as they occur.

Some of the primary areas include only a small number of stores, and in these areas it is feasible to prepare locally a complete list of establishments in the area and to be currently informed on births and deaths of establishments as they occur. For most of the primary areas, however, the stores involved are too numerous to make it feasible to list all stores locally and to follow all births and deaths of establish-

ments within the areas as they occur, and for these areas a method of subsampling from the selected primary areas will be used involving a further application of an area sampling procedure.

This area sampling procedure consists of designating small areas within the selected primary areas. The small areas will represent the stores not included on the large store list and the maintenance of up to date listings of stores in these areas will reflect the turnover in business establishments due to births and deaths.

The problem in designating small areas to be sampled within the selected primary areas is to locate and utilize the best available information and facilities for defining efficient subsampling units. The types of sampling principles already described as well as additional ones will be utilized in arriving at an efficient subsampling design.

For cities of 25,000 population or over the detailed Sanborn Insurance Maps referred to earlier provide an excellent mapping source that can be used in defining subsampling units. These maps are brought up to date at frequent intervals—about once a year, and show the locations of most stores within blocks in all such cities. Using these maps, it is possible to approximate the number of stores in each block and draw a sample of blocks stratified by size. An alternative possibility which will be followed sometimes is to designate small areas which have, according to the map, one, two, three, or as many stores as we desire, to be used as subsampling units. Such areas can be designated for inclusion in the sample by giving each block a chance of coming into the sample which is proportionate to its size in terms of number of stores, and then each sample block so selected can be segmented into areas of approximately equal numbers of stores. One of these areas is chosen at random from each sample block for inclusion in the sample.[7]

The methods used in sampling these small areas are such that any errors in the number of stores shown on the map will not bias the sample, although the poorer the counts made from the maps, the larger will be the sampling variance. Actually, the counts of stores from the maps have been found to be of such a quality that substantial reductions in sampling error have been made possible through their use.

In cities of 2,500 to 25,000 population, the Sanborn maps are not being utilized. The particular cities of this size to be included in the sample are obtained by a subsampling process within the selected counties and the 1940 Census returns are used in designating the ones

[7] For the theory involved in sampling with probability proportionate to size, see, Morris H. Hansen and William N. Hurwitz, "On the Theory of Sampling from Finite Populations," *Annals of Mathematical Statistics*, December, 1943.

to be included in the sample. Each city has a chance of selection proportionate to the number of stores in that city according to the 1940 Census. In most of the counties, however, there are only a few if any cities of this class, and all of them in the county may be included in the sample. For the cities of this size that are in the sample a straight sample of city blocks is being drawn for enumeration outside of the downtown area of such city, and a complete listing of store addresses will be made for the downtown areas and a subsample of stores or of small areas will be designated for coverage based on the information obtained in this listing.

To obtain a sample of the cities under 2,500 population and of other rural areas, minor civil divisions are used as sampling units. For these, 1940 Census information on the number of stores in each area is not available, and rough estimates of number of stores in each minor civil division are computed. These estimates are based on the 1940 population of the areas and on the proportion of that population living inside and outside of towns and villages. Then the actual areas to be included in the sample are selected with probability proportionate to this measure of size, or are stratified on the basis of this measure of size as was indicated above for the selection of blocks in the larger cities.

For the cities of all sizes, once a set of small areas is drawn into the sample all stores in those areas are listed. The enumerator is given a map showing the boundaries of the areas and is instructed to canvass the areas and list every store included therein. The advantage of the area approach is that it gives all stores not on the large store list a chance of being drawn into the sample, including both new establishments and ones inadvertently omitted from the large store list.

METHOD OF ESTIMATION FROM THE SAMPLE

A fundamental part of a survey design is the method of estimation from the sample. Equally fundamental is the method of evaluating the reliability of the estimates. The main aspects of the method of estimation and for evaluating contributions to the sampling error for samples of the type here outlined is presented in another paper that describes the methods and principles of estimation for the labor force sample.[8] The data from the 1940 Census of Retail Trade will be effectively utilized in preparing estimates from the sample, and will be utilized in such a way as to increase substantially the reliability of the

[8] See reference cited in footnote 7; and see also reference cited in footnote 6.

sample estimates through the use of ratio or regression types of estimates.

SUMMARY REMARKS

There will necessarily be variations introduced into the design as the plans are developed and put into operation, but the main principles followed will be those described above. Once the work in the field has progressed over a period of time further variations in the design will be introduced. For example, procedures may be developed through experience for handling an extensive list of large stores expeditiously in which event the large store list will be extended considerably. Again, the number of primary areas covered may be extended to make possible regional or State estimates as well as more detailed and accurate national estimates. At the same time, changes will be introduced as called for through further study of the costs involved, the sources of contributions to the sampling error, and the problems arising in carrying out the survey in the office and in the field.

ACTUARIAL ANALYSIS OF THE OPERATING LIFE OF B-29 AIRCRAFT ENGINES

OSCAR L. ALTMAN AND CHARLES G. GOOR*
Washington, D. C.

This study describes the actuarial methods used to determine the probable service life of the R-3350 engines used in B-29 aircraft. The use of actuarial methods to determine the expectation of life of engines under actual operating conditions is clearly superior to that of using averages based on the flying hours attained solely by engines removed for overhaul. This superiority is greatest when the engine "age" distribution or operating conditions are changing.

Though these methods and procedures were used in planning R-3350 engine requirements, production and shipping schedules, they are generally applicable in determining the number of replacements necessary to maintain a predetermined schedule of operation and can also be used in setting rates of depreciation.

This article presents the number of hours R-3350 engines ran before removal for overhaul, measures the improvement in successive models of engines, determines the difference in life between new and overhauled engines, and shows the effect of different operating uses upon engine life.

IT WAS necessary, very early in the B-29 aircraft program, to estimate the expectation of life of R-3350 engines prior to removal for overhaul, and to develop sensitive measures of actual operating life in order to adjust these estimated expectations of life (planning rates) to experience. The expected number of operating or flying hours prior to overhaul was a major element in determining the number of engines required each month to support the B-29 training program in the United States and to make possible the planned number of combat and other tactical missions against the enemy. The life of these engines was also a major element in determining the number of flyable engines that would be available month by month to support the B-29 program. Thus, decisions with respect to R-3350 engine requirements and availability vitally affected a whole series of inter-dependent variables, including (1) the ultimate size and the rate of expansion of the B-29 striking force, (2) the number and length of missions of B-29 aircraft, and (3) the required amount and rate of expansion of R-3350 engine production

* The procedures and data described in this article were developed by Lt. Col. Altman and Mr. Goor while they were attached to the Analysis and Plans Branch, A-4, Headquarters Army Air Forces. The authors wish to thank three other members of the Branch for their contributions: Mrs. Nina W. Levin and Miss Catherine Gordon for their assistance in the preparation and analysis of the original data; and Lt. Col. Harold Lamberton for the preparation of all charts involved in these studies.

and overhaul facilities. Accurate decisions with respect to the expansion of production and overhaul facilities were particularly important. On the one hand, the relatively long time required to obtain, fabricate and assemble the materials and components made engine production schedules relatively inflexible. On the other hand, expansion of facilities required large amounts of critical tools, materials and labor. For these reasons, the actual and planned operating life of R-3350 engines under combat and training conditions was a particularly important determinant of the entire B-29 aircraft program and therefore of the long-range air attack on Japan.

Engines for B-29 aircraft were critically short in 1944 and the first half of 1945, and the supply was extremely tight even at the end of the war with Japan. For many months air transportation was used to rush R-3350 engines to combat units overseas and then back to the United States for overhaul. This was no small feat since the great bulk and weight of these engines made it impossible to carry more than two R-3350 engines in the largest Army Air Forces transport.

All these factors created great pressure for accurate measurement of engine life in order to make possible the most effective supply and distribution planning and control, and led to the establishment, for the first time in the Army Air Forces, of the record-keeping and field reporting of the data required for the accurate determination of engine life by actuarial methods.

The fundamental purpose of this actuarial analysis of R-3350 engine life was to measure the average length of life of these engines, as well as their average future lifetime (expectation of life) after any specified period of flying time. The basic data used were the ratios of the number of hours flown by each engine removed for overhaul, to the number of hours flown by *all* engines, including those not yet removed for overhaul. Until actuarial methods were adopted, engine life prior to overhaul was measured solely by the average number of hours flown by engines removed for overhaul.

The latter method of necessity incorrectly showed low average engine life for many months after the introduction of a new type of engine for two reasons: (1) When *all* engines have flown a relatively small number of hours, the average number of hours flown by *removed* engines must also be relatively small. Thus, if no installed engine in a group of operating aircraft has accumulated more than one hundred hours of flying time, any engine removed for overhaul from this group would have a maximum of one hundred hours, and the average number of hours on *all* engines removed would be considerably less than one hun-

dred hours. As the surviving engines of this group accumulate more hours, both the maximum and the average flying hours on removed engines would increase. (2) The "age" distribution of installed engines would be affected by the introduction of large numbers of new aircraft into the group. These new aircraft would have engines that have flown only a small number of hours. While the introduction of new aircraft would not affect the rise of the *maximum* number of hours at removal of engines in this expanded group, it would retard the increase in the *average* number of hours at removal. In short, the method of measuring the number of hours flown by engines prior to overhaul solely on the basis of engines removed for overhaul always yields incorrect results under changing conditions of warfare because the number and "age" distribution of installed engines is constantly changing. This method can yield accurate results only after the "age" distribution of installed engines has been stabilized, and then only if operating conditions continue unchanged. Until the "age" distribution becomes standardized, which may take a year or more under usual rates of build-up of the striking force to peak strength, the average number of hours flown by removed engines prior to their removal would rise month by month. In the meantime, use of these averages as a planning rate for the future or as a measurement of performance in the past would be seriously in error and hopelessly inflate engine requirements.

Actuarial methods make it possible, on the other hand, to calculate the expectation of life of engines, with substantial accuracy, both while the "age" distribution is changing and after it has stabilized.

BASIC DATA

The data were collected in greater detail as the actuarial method of analysis gradually gained more and more support. By the spring of 1945 the following data were being collected by a training air force in the United States each month:

1. Number of installed engines:
   a. On hand, beginning of period
   b. Arrived during period
   c. Lost during period
   d. Departed during period
   e. On hand, end of period
2. Number of engines removed for overhaul:
   a. For accident, modification or other arbitrary reason
   b. For "normal" engine failure

Each major using organization in the United States and overseas

Line No.	Flying Hours Since Manufacture or Last Overhaul	Number of Installed Engines: On Hand at Beginning of Period	Number of Installed Engines: Arriving During Period	Number of Installed Engines: Lost During Period	Number of Installed Engines: Departing During Period	Number of Installed Engines: On Hand at End of Period	Engines Removed for Overhaul During Period: For Accident, Combat Damage, Modification, or Other Arbitrary Reasons	Engines Removed for Overhaul During Period: For "Normal" Engine Failure	Engines Exposed to Contingency of Failure	Per Cent Removed for Overhaul (Removal Rate)	Smoothed Data: Per Cent Removed for Overhaul	Smoothed Data: Per Cent of Original Engines Surviving	Smoothed Data: Expectation of Life (Hours)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1)	0											100.0	310
(2)	1– 20	153	241	2	27	94	3	7	384.0	1.8	.8	99.2	293
(3)	21– 40	56	258	—	39	113	—	5	592.0	.8	1.0	98.2	276
(4)	41– 60	21	87	1	28	57	1	8	678.5	1.2	1.2	97.0	259
(5)	61– 80	41	58	5	46	44	1	6	682.5	.9	1.4	95.6	243
(6)	81–100	65	26	1	20	27	8	14	695.5	2.0	1.6	94.1	226
(7)	101–120	69	19	4	20	30	2	17	715.0	2.4	1.9	92.3	210
(8)	121–140	49	24	—	19	30	6	17	723.0	2.4	2.3	90.2	195
(9)	141–160	27	24	1	12	31	—	25	718.5	3.5	2.7	87.8	180
(10)	161–180	30	21	1	18	51	7	36	684.0	5.3	3.2	85.0	166
(11)	181–200	42	15	1	24	22	7	40	636.5	6.3	3.9	81.7	152
(12)	201–220	37	6	—	14	45	11	53	584.5	9.1	4.6	77.9	139
(13)	221–240	16	7	—	21	40	12	33	493.0	6.7	5.4	73.7	126
(14)	241–260	21	17	—	24	29	8	27	423.5	6.4	6.5	68.9	115
(15)	261–280	24	6	2	11	13	13	25	380.5	6.6	7.6	63.7	103
(16)	281–300	16	7	1	9	21	11	39	341.5	11.4	9.1	57.9	92
(17)	301–320	24	2	—	11	12	13	31	288.0	10.8	10.9	51.6	82
(18)	321–340	16	6	—	7	12	10	36	248.5	14.5	12.9	44.9	73
(19)	341–360	8	4	—	4	7	8	36	205.5	17.5	15.2	38.1	65
(20)	361–380	5	—	—	1	13	7	30	158.0	19.0	18.0	31.2	57
(21)	381–400	7	—	1	—	4	1	28	120.5	23.2	21.3	24.6	49
(22)	401–420	10	1	—	—	7	—	20	95.0	21.1	25.3	18.4	42
(23)	421–440	2	4	—	3	2	—	16	77.5	20.6	30.0	12.9	36
(24)	441–460	—	—	—	4	6	—	21	57.0	36.8	36.0	8.3	30
(25)	461–480	1	—	—	1	7	—	9	27.5	32.7	42.0	4.8	25
(26)	481–500	—	—	—	—	4	—	3	13.0	23.1	50.0	2.4	21
(27)	501–520	—	—	—	—	1	—	3	7.5	40.0	59.0	1.0	16
(28)	521–540	—	—	—	—	—	—	2	4.0	50.0	70.0	.3	10
(29)	541–560	—	—	—	—	—	—	1	2.0	50.0	84.0	—	—
(30)	561–580	—	—	—	—	—	—	1	1.0	100.0	100.0	—	—
	Total	740	833	20	363	722	129	559					

Note on method of computation: The method of computation is described in general terms in the text. This is illustrated below for line 21 of the above table:

The data in Columns 1–8 are given.

The entry 120.5 (Col. 9, line 21) equals (a) the sum of lines 21–30 in Column 8, or 104; plus (b) the net total of Column 6 minus Column 2 for lines 22–30, or 14; plus (c) one-half the net total of Column 6 minus Column 2 for line 21, or −1.5; plus (d) the net total of Column 4 plus Column 5 plus Column 7 minus Column 3 for lines 22–30, or 3; plus (e) one-half the net total of Column 4 plus Column 5 plus Column 7 minus Column 3 for line 21, or 1.

The entry 23.2 (Col. 10, line 21) equals 28 (Col. 8, line 21) divided by 120.5 (Col. 9, line 21) and multiplied by 100.

The entry 21.3 (Col. 11, line 21) was read from the smooth curve fitted to the crude removal rates, as described in the text.

The entry 24.6 (Col. 12, line 21) equals (a) 100.0 minus 21.3 (Col. 11, line 21), or 78.7; multiplied (b) by 31.2 (Col. 12, line 20); divided (c) by 100.

The entry 49 (Col. 12, line 21) equals (a) the sum of Column 12 for lines 22–30, or 48.1; multiplied (b) by 20, to give 96.2; divided (c) by 24.6 (Col. 12, line 21), to give 39; plus (d) one-half of 20 hours.

was required to report these data in 20-hour flying time groups, with time measured as the number of hours flown since manufacture or last overhaul. The data were reported separately for new and overhauled engines and for each model of R-3350 engine. (The various engine models represented various stages in the development and improvement of the engine, beginning with the unmodified R-3350-23A carburetor type engine used in training and in early operations from India, followed by the modified R-3350-23A carburetor type and then later by the R-3350-57 fuel injection type.)

Collection of the data in this detail has made it possible to compare the relative performance of the various engines models, the performance of the same model in different theatres of action and under different operating conditions, the relative performance of new versus overhauled engines and the change in performance of any one model through time.

METHOD OF ANALYSIS

From these data were calculated, in terms of 20-hour flying time groups, the number of engines exposed to the contingency of failure, the crude and smoothed per cent removed for overhaul, the per cent of original engines surviving and the expectation of life. A sample calculation is shown in Table 1. The general method of computation may be described as follows with references to the columns in Table 1, assuming that frequency distributions by 20-hour flying time groups were reported for each of the following items:

(1) Number of engines removed for "normal" engine failure (col. 8).
(2) Number of engines removed for accident, combat damage, modification or other arbitrary reasons (col. 7).
(3) Number of installed engines on hand at the beginning of the period (col. 2).
(4) Number of installed engines on hand at the end of the period (col. 6).
(5) Number of installed engines on aircraft that arrived within reporting jurisdiction during period (col. 3).
(6) Number of installed engines on aircraft lost during period (col. 4).
(7) Number of installed engines on aircraft that departed from reporting jurisdiction during period (col. 5).

The exposure in any hour group X, is as follows:

$$[\text{col. } 9]_x = \sum_{i=0}^{\infty} [\text{col. } 8]_{x+i} + \sum_{i=1}^{\infty} [\text{col. } 6 - \text{col. } 2]_{x+i} + \tfrac{1}{2}[\text{col. } 6 - \text{col. } 2]_x$$

$$+ \sum_{t=1}^{\infty} [\text{col. } 4 + \text{col. } 5 + \text{col. } 7 - \text{col. } 3]_{x+t}$$

$$+ \tfrac{1}{2}[\text{col. } 4 + \text{col. } 5 + \text{col. } 7 - \text{col. } 3]_x$$

where the column numbers refer to the items in Table 1, Col. 9 being the "Engines Exposed to Contingency of Failure."[1]

The crude removal rates were then computed as a percentage by dividing the number of engines removed for normal engine failure in each hour group by the exposure in that group.

Analysis of a large number of these crude engine removal rate curves indicated that they had similar trends. When these crude rates were plotted on semi-logarithmic paper, they approximated straight lines. As a simple and quick method of graduation was desired, the data were plotted on semi-logarithmic paper with hours on the arithmetic scale. A straight line was then fitted by inspection. Chart 1 illustrates the results obtained by this procedure.

The percentage of the original number of engines surviving to the end of an hour group was computed by subtracting the per cent removed in that hour group from 100.0 per cent and by applying the remainder divided by 100 to the percentage of engines that survived to the beginning of that hour group.

The average length of life or the average future lifetime was then obtained by using the standard procedure for calculating the "complete

[1] This formula is an adaptation of the general "exposed to risk" formulae used in the calculation of mortality tables for human lives. The following are the basic elements in the construction of the formula:

(1) Every engine that reached an hour group greater than X during the period of observation must have passed through the hour group X during the period of observation unless it began the period or entered the experience during the period, at an hour group greater than X.

(2) Therefore, for any hour group X, the formula counts the number of engines that passed through that hour group during the period of observation as the sum of (a) the engines removed for "normal" engine failure at hours greater than X, during the period of observation; plus (b) the engines deducted from the experience because of "other" reasons for removal, loss while installed in aircraft or departing while installed in aircraft, at hours greater than X, during the period of observation; plus (c) the engines still installed at the end of the period of observation with hours greater than X.

(3) From the total in (2), above, are subtracted the engines that could not possibly have passed through the hour group X during the period of observation because they (a) started the period of observation as installed engines at hours greater than X; and (b) arrived installed in aircraft during the period of observation at hours greater than X.

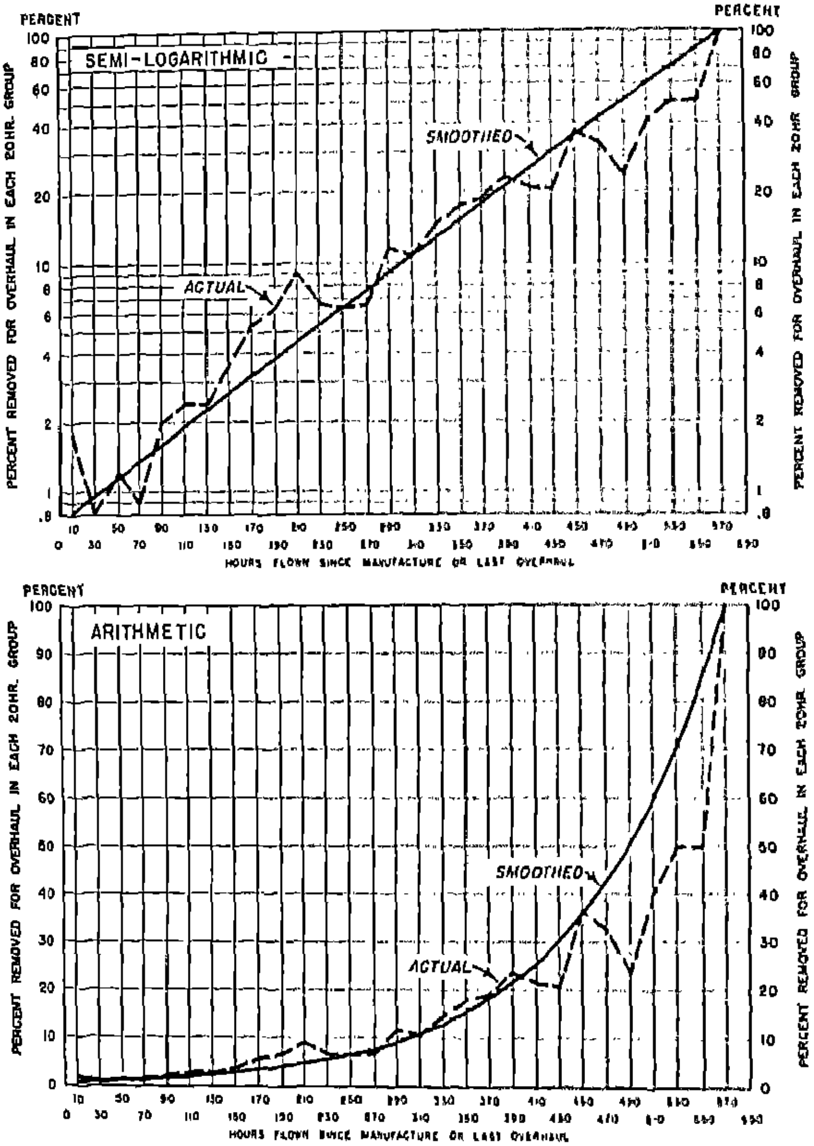
(4) In hour group X, a correction of ½ is made for each engine in that hour group, except engines removed for "normal" engine failure, on the assumption that they are uniformly distributed through the hour group and are therefore exposed to failure for only ½ the time.

(5) Each engine removed for "normal" engine failure in hour group X is counted as one because these removals during the hour group are considered as exposed to the end of the hour group.

For a more complete discussion of the general "exposed to risk" formulae, the reader is referred to H. S. Beers, "Notes on Exposure Formulae," *Transactions of the Actuarial Society of America*, Vol XLV, Page 41, and following discussion, May 1944.

CHART 1

OPERATING LIFE OF R-3350-23 ENGINES
ACTUAL AND SMOOTHED REMOVAL RATES
TRAINING OPERATIONS IN THE UNITED STATES, 1945



expectation of life." Specifically, this was accomplished for the end of any hour group X by (1) obtaining the sum of all survivorship per centages for all hour groups higher than X; (2) multiplying this sum by 20 hours to convert from the hour groups to the number of hours; (3) dividing this by the survivorship percentage of group X; and (4) adding one-half the hour interval to the result.

FINDINGS

Improvement in expectation of life of new R-3350 engines.—Sufficiently reliable information was collected to measure the effect of improvements made in successive models of the R-3350 engine upon the number of hours flown prior to their removal for first overhaul. Early combat operations in India, during the summer and fall of 1944, were conducted with B-29 aircraft equipped with unmodified R-3350-23A engines, which had an expectation of life of 163 hours prior to first overhaul. In February 1945, when the B-29 aircraft were equipped with -23A engines which had been modified and improved with more efficient cooling devices, the expectation of life of new engines rose to 280 hours, an increase of 117 hours per engine. The length of service was also increased substantially, as follows:

PER CENT OF NEW R-3350-23A (CARBURETOR TYPE) ENGINES SURVIVING TO SELECTED FLYING HOURS IN OPERATIONS FROM INDIA

Model	100 Hours	200 Hours	300 Hours	400 Hours
Unmodified	80.0	33.0	0.2	0
Modified	95.3	81.5	47.3	5.7

In February and March 1945, while the expectation of life prior to removal for first overhaul of the modified -23A engine was 280 hours in aircraft operating from India, it was 304 hours in aircraft operating from the Marianas, due largely to easier flying conditions in the islands. Later measurements indicated that the 304 hours would probably have been approximately 316 hours with more refined data.

The next model engine put into use was the fuel injection-equipped R-3350-57. Information is not available on the expectation of life of this engine prior to removal for the first overhaul in overseas operations. In the United States, however, the expectation of life of fuel injection engines was 329 hours compared with 310 hours for carburetor types engines. The increase in the expectation of life of 19 hours seems to indicate that the expectation of life of new fuel injection engines is some-

what better than of carburetor engines. The operating survival distribution was also lengthened somewhat, as follows:

PER CENT OF NEW R-3350 CARBURETOR AND FUEL INJECTION ENGINES SURVIVING TO SELECTED HOURS IN OPERATIONS IN THE UNITED STATES

Model	100 Hours	200 Hours	300 Hours	400 Hours	500 Hours
Carburetor (-23A)	94.1	81.7	57.0	24.0	2.4
Fuel Injection (-57)	96.0	84.3	63.3	31.2	4.9

It should be noted, however, that the fuel injection engines are safer and easier to operate, and the changeover was primariy planned for these reasons.

Expectation of life of new and overhauled engines.—The number of hours flown by new engines prior to removal for first overhaul is substantially greater than that flown by overhauled engines prior to second or later overhaul, despite the fact that during overhaul engines are completely disassembled, all worn parts replaced and the re-assembled engine thoroughly block-tested. Experience in the Marianas showed that the expectation of life was 316 hours for new engines and 270 hours for overhauled engines. Experience in training operations in the United States showed that the expectation of life was 310 hours for new, and 249 hours for overhauled engines. Thus the expectation of life of overhauled engines was 15 per cent lower in combat and 20 per cent lower in training than that of new engines.

The operating life of new engines before removal for overhaul was substantially longer than that of overhauled engines, both in the United States and overseas in the Marianas, as shown in the following table:

PER CENT OF NEW AND OVERHAULED R-3350-23A ENGINES SURVIVING TO SELECTED FLYING HOURS OVERSEAS AND IN THE UNITED STATES

Location and Overhaul Status	100 Hours	200 Hours	300 Hours	400 Hours	500 Hours
Overseas—combat					
New	96.8	87.5	62.7	19.6	0
Overhauled	92.5	75.5	43.4	8.8	0
United States—training					
New	94.1	81.7	57.0	24.0	2.4
Overhauled	87.0	65.4	30.4	10.4	0.5

It will be noticed that both new and overhauled modified carburetor engines in combat in the Marianas had greater expectations of life

than the respective groups in training in the United States. Ordinarily it would be expected that engines used in combat would not last as long as those in training. In this case, however, the situation was reversed. The very long range B-29's could climb very gradually after the take-off from the Marianas, thus saving the gasoline and avoiding the engine strain required to gain altitude quickly. In the United States, the airplanes were flown to high altitudes more quickly and, in addition, were probably put through extensive flying maneuvers as part of training operations. Furthermore, because of the extremely long distances to the target, the planes in the overseas group were flown "in combat" only a very small proportion of the total flight time to and from the target.

On the other hand, both the unmodified and the modified carburetor engines showed a much lower expectation of life in India than in the United States. The differences for new engines of both types are shown in the following table:

	Unmodified	*Modified*
India	163 hours	280 hours
United States	221 hours	310 hours

This was undoubtedly caused by the rugged flying conditions from India to enemy targets, which included rapid climb to get over the "Hump," and landing for refueling in China, followed by another take-off and another rapid climb to reach high altitudes over enemy held territory.

Use of expectation of life and survivorship percentages.—Requirements for spare engines to replace those removed for overhaul were determined by applying the estimated expectation of engine life to the planned schedule of hours to be flown. The planning factor for the unmodified R-3350-23A, adopted long before any B-29's were put in quantity operation, was 200 hours. Actual operations proved that this global planning rate was somewhat high. This rate resulted in an understatement of requirements in India, where the actual expectation of life was 163 hours. This understatement was compensated in part by an overstatement of requirements in the United States, where the actual proved to be 221 hours.

The global planning factor for the modified R-3350-23A, adopted long before these engines were put in quantity operation, was 300 hours. This rate proved to be remarkably accurate, and had the virtue of slightly overstating requirements for the Marianas and slightly understating those in the United States.

By the summer of 1945 it was clear that the planning rate of 300

hours would have to be revised downward for 1946 since the increasing proportion of overhauled to new engines would result in a lower expectation of life for all carburetor type engines, which would not be offset by the somewhat better performance of the fuel injection engines which were scheduled gradually to replace the carburetor engine.

For short-run supply planning and distribution for individual groups or areas, the distribution of survivorship percentages by 20 hour flying time groups (as shown in Table 2 and summarized in Chart 2), should be used instead of an expectation of life. The survivorship percentages

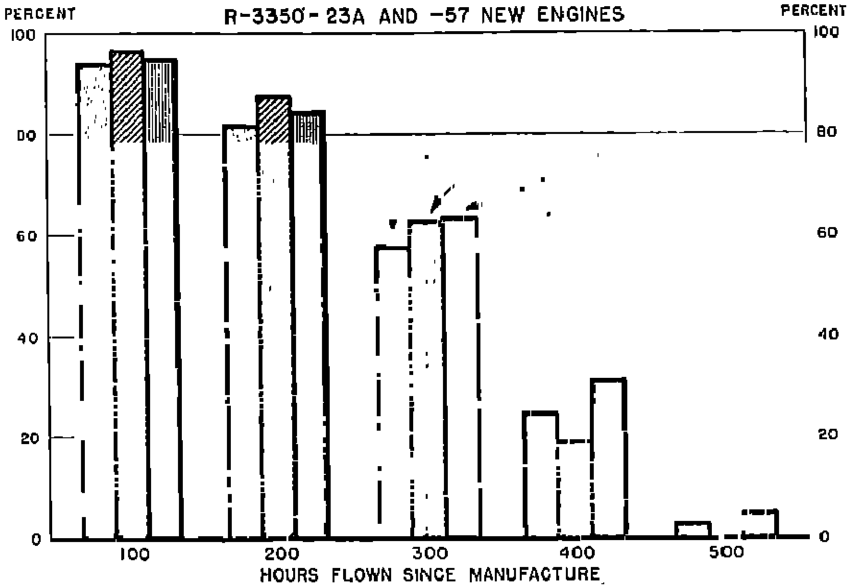
TABLE 2

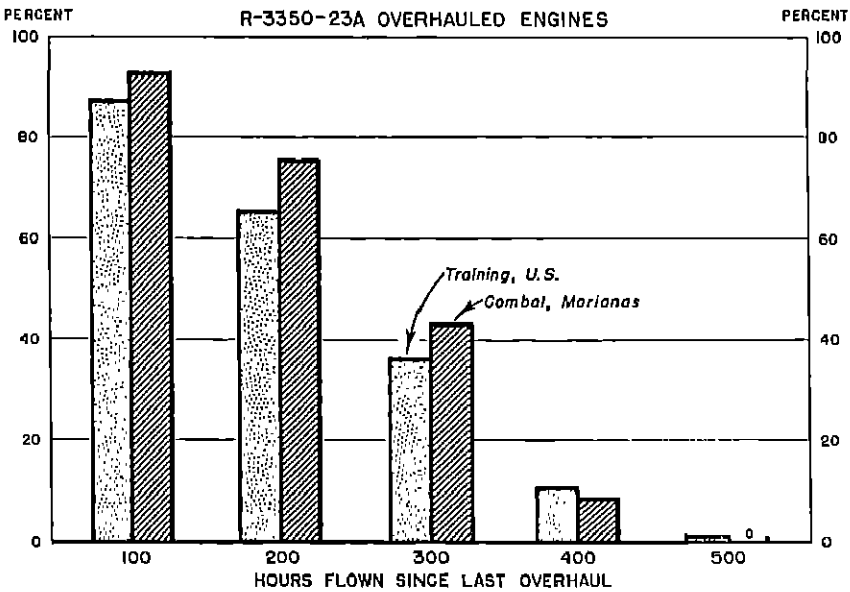
PER CENT OF ORIGINAL ENGINES SURVIVING AT EACH 20 HOURS STUDY AS OF 31 JULY 1945

Number of Hours	Fuel Injection	Carburetor			
	Training, U. S.	Training, U. S.		Combat, Marianas	
	New	New	Overhauled	New	Overhauled
0	100.0	100.0	100.0	100.0	100.0
20	99.3	99.2	98.0	99.0	99.0
40	98.5	98.2	95.7	99.1	97.8
60	97.5	97.0	93.1	98.5	96.3
80	96.3	95.0	90.2	97.7	94.0
100	95.0	94.1	87.0	96.8	92.5
120	93.5	92.8	83.4	95.6	90.1
140	91.7	90.2	79.5	94.2	87.2
160	89.6	87.8	75.2	92.4	83.9
180	87.2	85.0	70.5	90.2	80.0
200	84.3	81.7	66.4	87.5	75.5
220	81.1	77.0	60.0	84.2	70.4
240	77.4	73.7	54.4	80.2	64.5
260	73.2	68.9	48.5	75.3	58.0
280	68.5	63.7	42.4	69.4	50.9
300	63.3	57.0	36.4	62.7	43.4
320	57.6	51.6	30.5	55.0	35.6
340	51.4	44.0	24.7	46.5	27.8
360	44.8	38.1	19.3	37.2	20.6
380	38.1	31.2	14.5	27.9	14.0
400	31.2	24.6	10.4	19.0	8.8
420	24.9	18.4	7.0	11.4	4.8
440	18.4	12.9	4.3	5.7	2.1
460	12.0	8.3	2.4	2.1	.7
480	8.4	4.8	1.2	.4	.1
500	4.9	2.4	.5	—	—
520	2.4	1.0	.2	—	—
540	1.0	.3	.1	—	—
560	.6	.1	—	—	—
580	.1	—	—	—	—

CHART 2

PER CENT OF ORIGINAL ENGINES SURVIVING TO EACH 100 HOURS BY STATUS AND ACTIVITY





should be converted into a table of probabilities of removal within the number of flying hours assigned for the period; and these probabilities should be applied to the flying time "age" distribution of the engines installed in aircraft at the beginning of the period to determine "normal" engine requirements.

TABLE 3

AVERAGE HOURS ON REMOVALS COMPARED WITH EXPECTATION OF LIFE OF R-3350-23A ENGINES IN COMBAT IN THE MARIANAS

Date	Expectation of Life	Average Hours on Removals
All engines		
30 Nov. 1944	—	91
31 Dec. 1944	—	133
31 Jan. 1945	—	151
20 Feb. 1945	271[1]	159
28 Feb. 1945	—	161
31 Mar. 1945	301	192
30 Apr. 1945[2]	310	234
31 May 1945	314	259
30 June 1945	308	268
31 July 1945	309	272
New Engines		
31 Mar. 1945	304	195
30 Apr. 1945[2]	310	236
31 May 1945	316	263
30 June 1945	316	273
31 July 1945	316	279
Overhauled engines		
31 Mar. 1945	289	136
30 Apr. 1945[2]	296	206
31 May 1945	294	215
30 June 1945	284	212
31 July 1945	270	205

[1] Losses of installed engines were omitted since data were not available for this calculation.

[2] Beginning 30 April 1945, removal rates were based on normal removals only. Removals directly attributable to accident or combat damage and engines removed for purposes of modification or model change were not used as removals.

Comparison of engine planning rates based on actuarial studies of all engines and studies of only removed engines.—Actual experience in the measurement and planning of R-3350 engine life prior to removal for overhaul demonstrated beyond doubt the superiority of the use of actuarial methods applied to the performance of both removed and installed engines over the use of averages based solely on the flying time of removed engines.

As was expected, the average number of hours flown by engines *removed for overhaul* increased steadily month after month. This upward

trend did not represent an increase in performance, however, but rather a change in the "age" distribution of installed engines. The average number of hours on engines removed from aircraft operating from the Marianas was 91 in November 1944, 161 in February 1945, and 272 in July 1945. (See Table 3.) Thus, even after nine months of operations, the average number of hours on removed engines had not yet attained the expectation of life of 316 hours to which it is asymptotic because the continuous and rapid increase of aircraft diluted the engine "age" distribution with new engines. Similar results were observed with respect to engines operating in the United States.

CONCLUSION

The use of actuarial methods to determine the expectation of life of engines under actual operating conditions is clearly superior to that of using averages based on the flying hours attained solely by engines removed for overhaul. This superiority is greatest when the engine "age" distribution or operating conditions are changing. The need for actuarial methods is most pressing when (1) costs are high and production, transportation or dollar resources are limited; (2) expectation of life of the engine is high, since it will then take many months, even under stable operating conditions, for the installed engine "age" distribution to become stabilized; and (3) the build-up of the total number of installed engines to its peak takes many months.

SYSTEMATIC SAMPLING AND ITS RELATION TO OTHER SAMPLING DESIGNS*

LILLIAN H. MADOW

This paper presents a simple introduction to the theory of systematic sampling. Its relation to random sampling, stratified random sampling and cluster sampling is shown by means of a numerical example. Some advice is given on the circumstances under which systematic sampling may be advantageously used.

IN DEVELOPING any given sample survey, many designs can yield sufficiently accurate estimates, if sufficiently large expenditures are made. The problem then is to select from among the various possible designs the most efficient one, i.e., the one that obtains the most information for a given expenditure of resources. The earliest statistical theory of sampling designs gave results for the unrestricted random design whereby any desired degree of accuracy could be obtained by altering the size of sample. In actually collecting data, however, the unrestricted random design was in general impractical administratively, and therefore rarely used. Alternative designs were used, some for which theory was available and some for which the estimates of error could only be guessed.

One basic sampling technique which has been in use for many years but for which until recently no supporting theory existed is the systematic sampling design, which, roughly speaking, consists in selecting every kth element of the population, the first element being selected at random. The systematic sample was often used in preference to other designs for which theory was available, because it was administratively simple and it was supposed that the results would be better than or at least close to those that would have been obtained by unrestricted random selection. It was also thought that the systematic design might take advantage of possible hidden stratification in the elements of the population. More refined considerations of the merits of the systematic design over the unrestricted random or other designs was hardly possible, since the theory of the systematic design was not known.

The beginnings of a theory of the systematic sampling design have already been given.[1] The results of that paper are here presented in less technical form and with numerical calculations carried out on il-

* Based on a thesis submitted to the Graduate Faculty of the American University in partial fulfillment of the requirements for the degree of Master of Arts, June 1944.

[1] "On the Theory of Systematic Sampling, I" by William G. Madow and Lillian H. Madow, in the *Annals of Mathematical Statistics*, March 1944.

lustrative examples. It is to be hoped that, with the theory of systematic sampling available to give the error of estimate of this design and to make effective comparisons with alternative designs, more confidence will be placed in a decision to use or not to use the systematic design in any given survey.

Illustrative example. In sampling a human population for a given characteristic, say to estimate the number of owner-occupied dwelling units in a given area, the procedure usually resolves itself into first dividing the area into strata by a geographic or rational grouping, and then estimating the characteristic within the strata by taking either an unrestricted random or a systematic sample. To compare these two designs in this situation, we shall assume that one of our strata consists of the 20 blocks of census tract S-6, Boston, Massachusetts, as enumerated in the 1940 housing census. This tract was selected only for its convenience in size. The data are presented in Table 1.

TABLE 1

CENSUS TRACT S-6, BOSTON, MASSACHUSETTS, APRIL 1, 1940

Block number	*Number of owner-occupied dwelling units* x_i	x_i^2
1	6	36
2	7	49
3	1	1
4	11	121
5	23	529
6	14	196
7	7	49
8	8	64
9	5	25
10	3	9
11	6	36
12	21	441
13	8	64
14	8	64
15	6	36
16	4	16
17	3	9
18	8	64
19	4	16
20	7	49
Σ	160	1874

Since the variance of the estimate of the mean of a sample drawn from a stratified population depends only on the within-strata variance

and not on the between-strata variance, any reduction in total variance must come from a reduction of the within-strata variance. For the data in Table 1, we shall investigate whether the unrestricted random or the systematic design gives the smallest within-strata variance. For this purpose, we can consider the Boston census tract a population P consisting of N elements, $x_1, x_2, \cdots x_N$, $N=20$. The mean value is given by

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i,$$

and the variance is

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2,$$

where by σ^2, the variance of the stratum which we are considering a subpopulation, is meant the variance of a random sample of one element from the population of N elements.

Substituting from Table 1 in the formula for $\bar{x}$ and σ^2, we see that

$$\bar{x} = \frac{160}{20} = 8.0 \quad \text{and} \quad \sigma^2 = \frac{1874}{20} - 64.0 = 29.7.$$

The problem now is to take for each proposed design a sample of n elements of the possible N, where $N=20$, to make an estimate, $\bar{\bar{x}}$, of the population mean, $\bar{x}$, where $\bar{x}=8.0$, and to find the variance of the estimate about the mean.

The unrestricted random sampling design. In this design, the procedure is to select at random n elements from the N possible ones, without replacement. The exact procedure may be something like the following. First, the elements of the population, if not already in some order, should be numbered from 1 to N. Then, if N is, say, 20, and it is desired to take a sample of 4 elements, a starting point in a 2-digit column of a table of random numbers is selected at random. If the random number is, say 16, then the element in the population numbered 16 is selected for the sample. If a number in the table is greater than 20, it may be skipped. This procedure is continued until 4 elements have been selected. If a second number 10 occurs in the table of random numbers, then it must be skipped, if the sampling design is to be without replacement. Other random procedures could also be defined.

For the design consisting of random sampling without replacement, the estimate of the population mean is given by

$$\bar{\bar{x}} = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

where $\bar{\bar{x}}$ is the arithmetic mean of the results of the n drawings, and x_i is the result of the ith drawing. This is an unbiased estimate of $\bar{x}$, and the variance of the estimate is given by

$$\sigma_{\bar{x}}^2 = \frac{N - n}{N - 1} \frac{\sigma^2}{n},$$

where N, n, and σ^2 have already been defined.

For the example given in Table 1, suppose we consider the sampling procedure that consists of selecting 4 blocks at random from the set of 20 blocks. Then, in practice, we calculate $\bar{\bar{x}}$ for the sample in order to estimate $\bar{x}$. Substituting in the above formula for $\sigma_{\bar{x}}^2$, we then have

$$\sigma_{\bar{x}}^2 = \frac{20 - 4}{20 - 1} \cdot \frac{29.7}{4} = 6.2526.$$

It is readily seen from the general formula, that the variance of this design depends only on the population variance, the size of n, and the relative sizes of n and N. To use the unrestricted random design effectively, then, it is necessary to have some rough idea of the variability of the elements in the population in order to estimate σ^2, and it is necessary to have a prelisting of the population so that a random design can be associated with the elements in the population. It is not necessary, however, to have any further knowledge about the characteristic of the population to be measured. It is not necessary to know whether the characteristic to be measured tends to cluster in groups or whether it is widely scattered throughout the population, or whether there is a serial trend in the characteristic. In collecting social or economic data, it would be unnecessary to know anything about the areal, color, or income level distribution of a population, if the design to be used is the unrestricted random one. The requirement of having a prelisting of the population that is up-to-date is one that is very often difficult to meet. If the prelisting is out of date, a random sample based on that prelisting will be biased.

The systematic sampling design. The theory of the systematic sampling design was developed directly in Reference 2, but it can also be derived in a simple manner from the theory of the cluster sampling design, since the systematic sampling design is cluster sampling in which the elements within clusters are arranged somewhat differently than usual.

In the usual type of cluster sampling, the procedure is to divide the population into classes or clusters, taking the elements of the popula-

tion in their natural order, i.e., the elements of each possible cluster are contiguous elements of the population. The sample then consists of several of the possible clusters, and there may be subsampling within clusters.

The theory of the systematic sampling design follows from the simplest case of cluster sampling, i.e., the case in which only one cluster is sampled and there is no subsampling within the cluster.[2] The big difference is that in the systematic design, if there are k possible clusters, the elements within any one of the clusters are k elements apart in the population, and no two elements within any of the possible clusters are contiguous elements of the population.

For the simplest cluster sampling design, the procedure is to divide the population of N elements into k clusters of n elements each. Only one cluster is taken in the sample and there is no subsampling within the cluster. In this case, each of the possible k clusters has the probability $1/k$ of being the one that is selected. $N = kn$.

For this design, we define $\bar{x}_i = \dfrac{\sum_{j=1}^{n} x_{ij}}{n}$, where x_{ij} is the jth element of the ith cluster, and $\bar{x}_i$ is the mean of the ith cluster, $i = 1, \cdots, k$. The estimate of the population mean is given by $\bar{\bar{x}}$, where $\bar{\bar{x}} = \bar{x}_i$, if the ith possible sample is the one that is selected. This estimate is unbiased, and the variance is given by

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left[1 + (n-1)\rho\right],$$

where ρ is the intraclass correlation coefficient and is given by the equation

$$\rho\sigma^2 = \sigma_b^2 - \frac{\sigma_w^2}{n-1}.$$

In this expression, σ_b^2 is the variance between clusters and σ_w^2 is the variance within clusters, and they are defined by

$$\sigma_b^2 = \frac{\sum_{i=1}^{k} (\bar{x}_i - \bar{x})^2}{k} \quad \text{and} \quad \sigma_w^2 = \frac{\sum_{i=1}^{k} \sigma_i^2}{k},$$

[2] For detailed treatment of the cluster sampling design, see "On the Theory of Sampling from Finite Populations," by Morris H. Hansen and William N. Hurwitz, *Annals of Mathematical Statistics*, vol. 14, No. 4 (1943), pp. 333–362.

respectively. From the formula for $\sigma_{\bar{x}}^2$, it is now possible to see how much the variance depends on ρ, the intraclass correlation coefficient. It is easily shown that when $\rho = -1/N-1$, then the variance of the cluster sampling design is the same as that of the unrestricted random sampling design. Making this substitution for ρ, we have

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left[1 - \frac{n-1}{N-1}\right]$$
$$= \frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right],$$

the well-known expression for the variance of a random sample of n observations from N when the sampling is without replacement.

Moreover, it is readily observed that if ρ is less than $-1/N-1$, the variance of the cluster sampling design is less than that of the random sampling design, and if ρ is greater than $-1/N-1$, then the variance of the cluster sampling design is greater. Or, roughly speaking, if ρ is negative, cluster sampling yields a smaller variance than does unrestricted random sampling; if ρ is positive, the random sampling variance is less; and if ρ is approximately equal to 0, then the two designs yield approximately the same variance, although the variance of the cluster sampling design will be the greater one.

It is now possible to apply this formula to the Boston census tract we are using for illustration. Since we want $n=4$, the design will be to divide the population of 20 elements into 5 clusters of 4 elements each and to sample one of the clusters at random.

TABLE 2*

	C_1	C_2	C_3	C_4	C_5	
	6 7 1 11	23 14 7 8	5 3 6 21	8 8 6 4	3 8 4 7	
$\sum_{j=1}^{4} x_{ij}$	25	52	35	26	22	160
$\bar{x}_i$	6.25	13.0	8.75	6.5	5.5	
$\lvert\bar{x}_i - \bar{x}\rvert$	1.75	5.0	.75	1.5	2.5	
$(\bar{x}_i - \bar{x})^2$	3.0625	25.0	.5625	2.25	6.25	37.125

* The same data as in Table I, but arranged in 5 clusters of 4 elements each.

Therefore,

$$\sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^{5} (x_i - \bar{x})^2}{5} = \frac{37.125}{5} = 7.425.$$

To calculate ρ, we let

$$7.425 = \frac{20.7}{4} [1 + (4 - 1)\rho],$$

and we find that $\rho = 0$.

If we rearrange the data of the Boston census tract so that the elements within the possible clusters are 5 elements apart in the population, we have a systematic arrangement of the data. The basic calculations for this design are given in Table 3.

TABLE 3*

	K_1	K_2	K_3	K_4	K_5	
	6 14 6 4	7 7 21 3	1 8 8 8	11 5 8 4	23 3 6 7	
$\sum_{i=1}^{4} x_{ij}$	30	38	25	28	39	
$\bar{x}_i$	7.5	9.5	6.25	7.0	9.75	
$\lvert \bar{x}_i - \bar{x} \rvert$	.5	1.5	1.75	1.0	1.75	
$(\bar{x}_i - \bar{x})^2$	.25	2.25	3.0625	1.0	3.0625	9.625

* The data are the same as in Table 1, but arranged by taking 5 clusters of 4 elements each, the elements within each of the clusters being every 5th element of the population instead of being consecutively arranged within clusters. Contiguous elements of the population are to be found across clusters.

Based on the arrangement in Table 4, and using the formula for the variance of the cluster sampling design just given, we calculate

$$\sigma_{\bar{x}}^2 = \frac{9.625}{5} = 1.925.$$

To find ρ, we let $1.925 = \frac{20.7}{4} [1 + (4 - 1)\rho]$ and get $\rho = -.247$.

For this arrangement of the data, then, we see that there is a tremen-

dous gain in efficiency as a result of the systematic design, in this particular example.

In general, in the systematic design for selecting a sample of n elements from a population of $N = kn$ elements, the procedure for selecting the sample involves the following steps.

1. One of the numbers from 1 to k is selected at random.
2. If the number selected in Step 1 is i, then the ith element is selected for the sample.
3. Every kth element following the ith element is selected for the sample.

Let $\bar{x}$ now be the average of the n elements selected according to the above procedure. Then, $\bar{x}$ is unbiased, and has the variance

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} [1 + (n - 1)\rho],$$

where ρ, the intraclass correlation coefficient has the same definition as before in cluster sampling. In this case, however, the clusters instead of having been, shall we say, given by nature, have been formed by the sampling procedure given above. There are k clusters, the ith cluster corresponding to the selection of the number i in Step 1 above.

To understand the character of the systematic design, i.e., to determine for which populations the systematic design is advantageous, it is desirable to substitute for the intraclass correlation coefficient a sum of serial correlation coefficients. In the cluster sampling design it is frequently too lengthy a job to calculate the intraclass correlation coefficient which is so important in estimating the necessary sample size. In many cases, however, it is possible from the nature of the characteristic of the population to be estimated to guess at a reasonably good figure for this intraclass correlation coefficient. At least, it is possible to determine whether it is expected to be positive, negative, or zero, just from the rational considerations of the problem, and an understanding of what ρ measures. Similarly, for the systematic design, it is useful to think in terms of the serial correlation coefficients, even though in a particular problem it may be too lengthy a job actually to perform the calculations. In the systematic sampling design, a good estimate of the variance depends on a good estimate of the serial correlation coefficients. These coefficients measure the degree of dependence of successive elements in the population.

The serial correlation coefficient, using the circular definition, is defined to be

$$\rho_{k\delta} = \frac{\sum_{\nu=1}^{kn} (x_\nu - \bar{x})(x_{\nu+\delta k} - \bar{x})}{kn\sigma^2}, \quad \text{where,}$$

if $\nu+\delta k > kn$, then its value is taken to be $\nu+\delta k-kn$, and if $\delta=n/2$, then

$$\rho_{k(n/2)} = \frac{\sum_{\nu=1}^{k(n/2)} (x_\nu - \bar{x})\left(x_{\nu+k(n/2)} - \bar{x}\right)}{kn\sigma^2}.$$

When we substitute for ρ in terms of the $\rho_{k\delta}$ in the formula for the variance of the systematic design, we obtain

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left[1 + 2\sum_{\delta} \rho_{k\delta}\right],$$

where $\delta=1, \cdots, n/2$, when n is even, and $\delta=1, \cdots, (n-1)/2$, when n is odd.

Comparisons between the efficiency of the systematic and random sampling designs can readily be made by observing that the variance of the systematic design is equal to that of the unrestricted random design, when

$$\sum_{\delta} \rho_{k\delta} = -\frac{n-1}{2(N-1)},$$

where $\delta=1, \cdots, n/2$ if n is even, and $\delta=1, \cdots, (n-1)/2$ if n is odd. If we substitute this expression for $\sum_{\delta}\rho_{k\delta}$ in the expression for

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left[1 + 2\sum_{\delta}\rho_{k\delta}\right],$$

where the limits for the summation with respect to δ are the same as those given above, then the expression reduces to

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right],$$

which is the well known expression for the variance of a random sample of n out of a finite population of N, when the sampling is without replacement. Moreover, it is readily observed that when the sum of the serial correlation coefficients is less than $-\frac{n-1}{2(N-1)}$, then the variance

of the systematic sampling design is less than that of the random sampling design, and if the sum of the serial correlation coefficients is greater than $-\dfrac{n-1}{2(N-1)}$, then the variance of the systematic sampling design is greater.

It is not possible to give simple directions for the conditions under which the systematic design is most appropriate. Some statements may, however, be helpful. Administratively, the systematic design is simple. Clearly, it is simpler to select every kth element from a card file, a map, or a prelisting, than to check with a table of random numbers before selecting each element to be included in the sample. The chief danger in applying the systematic design occurs when the data have a periodic formation, and the sampling interval chosen is equal to the period of the data. Past experience will provide useful information in the prediction of the relative efficiencies of different designs. When evidence is obtained that the systematic design is effective for one size of sample, care must be taken not to assume that it is effective for even a relatively close size of sample. As an illustration of the dangers of the uncritical use of the systematic design, an analysis of some data on seedlings is presented. In addition to testing the use of the systematic design, comparisons for several sample sizes are also made with the unrestricted random and the stratified random sampling designs.

Before giving the calculations for the seedling data, it might be well to use the Boston census tract data again to indicate the calculations that are made in connection with studying the applicability of the systematic design to a set of data. The calculations are too voluminous to reproduce for the seedling data, and therefore only the results will be given for them. For the Boston census tract, however, it is possible to give all the calculations.

The basic data and calculations are given in Table 4.

We have already calculated for these data that $\bar{x}=8.0$ and $\sigma^2=29.7$. From Table 5 we can now calculate

$$\sigma^2\rho_{5,1} = \frac{1101}{20} - 64.0 = -8.95; \qquad \therefore \quad \rho_{5,1} = -.30135, \quad \text{and}$$

$$\frac{1}{2}\,(\sigma^2\rho_{5,2}) = \frac{1}{2}\left(\frac{1108}{20} - 64.0\right) = -2.05;$$

$$\therefore \ \frac{1}{2}\rho_{5,2} = -\ .06902 \qquad \therefore \ \sum_{k=1}^{n/2} \rho_{ki} = -\ .37037$$

Therefore, $\sigma_{\bar{x}}^2 = \dfrac{29.7}{4}\,[1+2(-.37037)] = 1.025$. This is the same figure that was obtained for the cluster design calculated on the basis of Table 4 when the elements were arranged in systematic order and only one

TABLE 4*

Block number	*Number of owner-occupied dwelling units* x_i	$x_i x_{i+5}$	$x_i x_{i+10}$	x_i^2
1	6	6·14 = 84	6·6 = 36	36
2	7	7·7 = 49	7·21 = 147	49
3	1	1·8 = 8	1·8 = 8	1
4	11	11·5 = 55	11·8 = 88	121
5	23	23·3 = 69	23·6 = 138	529
6	14	14·6 = 84	14·4 = 56	196
7	7	7·21 = 147	7·3 = 21	49
8	8	8·8 = 64	8·8 = 64	64
9	5	5·8 = 40	5·4 = 20	25
10	3	3·6 = 18	3·7 = 21	9
11	6	6·4 = 24	6·6 = 36	36
12	21	21·3 = 63	21·7 = 147	441
13	8	8·8 = 64	8·1 = 8	64
14	8	8·4 = 32	8·11 = 88	64
15	6	6·7 = 42	6·23 = 138	36
16	4	4·6 = 24	4·14 = 56	16
17	3	3·7 = 21	3·7 = 21	9
18	8	8·1 = 8	8·8 = 64	64
19	4	4·11 = 44	4·5 = 20	16
20	7	7·23 = 161	7·3 = 21	49
$\sum_{i=1}^{20}$	100	1101	1198	1874

* The same data as given in Table 1, with some additional columns of computation.

cluster of 4 elements was chosen in the sample. This agreement is in accordance with the theory.

Data on seedlings. Calculations of the same type were carried out on some seedling data collected by Floyd A. Johnson, Iowa State College. The data consist of 420 observations on seedlings in one bed of a forest nursery. The bed had 4 rows and was 420 feet long. Each observation represents one foot of bed-width, or the total of the four rows in that particular foot of bed. Some of the results are given in Table 5.

TABLE 5

n	k	$2\sum_{\delta} \rho_{k\delta}$*	σ_x^2
10	42	−.2097	4.213
21	20	−.1045	3.061
28	15	−.1781	2.148
30	14	−.4012	.693
42	10	−.1521	1.742
60	7	−.4250	.203
84	5	−.0120	1.220

* $\rho_{k\delta}$ is summed from $\delta=1$ to $\delta=n/2$ if n is even or to $n-1/2$ if n is odd.

For the same values of n as given in Table 5, calculations were made for the variance of the unrestricted random and the stratified random designs. The stratified random design was simple proportionate sampling within strata. The strata were of equal size and one element was selected from each stratum. The results are given in Table 6 together with a comparison of the relative efficiencies of the systematic and stratified random designs with respect to the unrestricted random design.

TABLE 6

n	σ_R^2	σ_{St}^2	σ_{Sys}^2	Eff_{St}	Eff_{Sys}
1	2	3	4	5	6
10	10.203	7.212	4.213	42.7	144.3
21	4.770	2.000	3.061	60.1	55.8
28	3.515	2.002	2.418	68.0	45.4
30	3.264	1.902	.693	71.6	371.0
42	2.250	1.292	1.742	74.8	29.7
60	1.506	.8183	.2631	84.0	472.4
84	1.004	.5052	1.220	98.7	−17.7

In Table 6, the first column represents the sample size, and columns 2, 3, and 4 represent the variances of the unrestricted random, the stratified random, and the systematic sampling designs, respectively. Columns 5 and 6 represent the relative gain in efficiency of the stratified random and systematic designs, respectively, over the unrestricted random design. The efficiency of the alternative designs was measured by dividing the variance of the unrestricted random design by that of the restricted design, and expressing the result as a per cent. By this measure, a restricted design has the same efficiency as the unrestricted random design, if the measure is 100 per cent. If the restricted design is more efficient than the unrestricted random design, then the measure will be greater than 100 per cent, and if the restricted design

is less efficient, then the measure will be less than 100 per cent. This measure of relative efficiency cannot be negative, of course, by definition.

Another way of looking at this measure of relative efficiency is in terms of sample size. For example, for $n=10$, Table 6 shows that for every element in a stratified sample, the unrestricted random sample requires about one and one-half elements to have the same efficiency. Similarly, for $n=10$, the random sample requires about two and one-half elements for every element of the systematic design to have the same efficiency.

Table 5 shows some interesting information about some characteristics of the systematic design. First, we notice that the value of $\sum_{\delta}\rho_{k\delta}$, although always negative, is extremely variable, and that it does not vary consistently with either n or k. The value of $\sum_{\delta}\rho_{k\delta}$, for any given value of n or k, depends not so much on the value of n or k as on the arrangement of the elements in the original population and the degree to which successive elements in the population are positively or negatively correlated. We also see that although, in general, the variance of the systematic design decreases with increasing size of n, this decrease is not uniform. For example, the variance at $n=30$ is much lower than that at $n=28$ or $n=42$. Again, the variance at $n=60$ is much lower than at $n=42$ or $n=84$.

From Table 6, we can compare the variances of the systematic design with two other designs for the same sample sizes. The variance of the systematic design for different values of n fluctuates, whereas the variances of the proportionate stratified random and unrestricted random designs behave in a stable manner. For the latter designs it can always be said that if the variance is known for a given n, the variance for a sample of $n+m$ will be smaller, no matter what the value of m. This statement does not hold for the systematic design.

The stratified random design shows a steady increase in efficiency over the unrestricted design for increasing values of n, as can be seen from Table 6. The efficiency of the systematic design as compared with the unrestricted random design, however, is erratic. The systematic design is more efficient than the stratified random design in the cases when $n=10$, 30, or 60, but it is less efficient than the stratified random design for the cases in which $n=21$, 28, 42, or 84. In all cases except for $n=84$, the systematic design is more efficient than the unrestricted random design, but for $n=84$ the relative efficiency is less than 100, being only 83.3 per cent.

For these data, the stratified random is more efficient than the unrestricted random design for all values of n, but it never attains the degree of efficiency that the systematic design attains. In the cases where the systematic design is more efficient than the stratified random design, the systematic design is about twice as efficient as the stratified random design, whereas in most of the cases in which the systematic design is less efficient than the stratified random design, the stratified design has only a slight gain over the systematic design.

CALCULATING THE GEOMETRIC MEAN FROM A LARGE AMOUNT OF DATA

ZENON SZATROWSKI
Northwestern University

This note presents a short procedure for calculating the geometric mean from a large amount of data. The method involves a) using a grouping of the frequency distribution such that the ratios of the class limits are equal, and b) an application of the "short method" for calculating the arithmetic average. By this procedure the calculation of the geometric mean takes more time than that of the arithmetic mean, only to the extent of getting the logarithms of four values and the antilogarithm of one value. Analogous procedure can be used for the harmonic and other means.

INTRODUCTION. Discussions in statistics texts leave the impression that the calculation of the geometric mean of a large number of measurements requires much more time than the computation of the arithmetic average. The procedures usually presented involve either a) the determination of the logarithms of each measurement, which is unnecessary; or b) the use of mid-points of equal intervals of the arithmetic values, which gives the resulting geometric mean an upward bias. The purpose of this note is to show that it is possible to calculate the geometric mean of a large number of measurements just about as easily as the arithmetic average. By this procedure the calculation of the geometric mean takes more time than that of the arithmetic mean, only to the extent of getting the logarithms of four values and the antilogarithm of one value. Furthermore, analogous procedure can be used in obtaining the harmonic and other means, which by definition, involve arithmetic averages of some functions of the original values. Grouping the frequency distribution with suitably chosen unequal intervals and the application of the so-called "short method" for calculating the arithmetic average are the basis of the procedure to be demonstrated.

The General Idea. The geometric and harmonic means are examples of a more general averaging process which may be described as follows: X, the original variable, is transformed by means of a relationship $Z=f(X)$. (For the geometric mean, $Z=\log X$). Then the arithmetic average of the Z's, $\bar{Z}$, is obtained and the desired average of the X's, $\hat{X}$, is merely the value of X corresponding to $\bar{Z}$. (For the geometric average, $\hat{X}=\text{antilog } \bar{Z}$. The idea of the proposed procedure is that, if a grouping method with equal class intervals on the Z scale is used,

then it is not necessary to apply the transformation $Z=f(X)$, for each measurement. It is sufficient to find the X values corresponding to the class limits on the Z scale. This yields unequal intervals on the X scale according to which the X values can be grouped.

Demonstration of Method. In the following, a short practical way of calculating the geometric average for a large amount of data is explained using the data in Table 1.

TABLE 1
THE DATA

515.8	354.7	201.3	300.7	237.3
252.0	223.8	270.2	370.0	364.0
301.0	334.3	349.0	300.7	204.3
288.9	318.8	353.7	417.1	350.5
386.0	189.0*	482.0	218.5	343.0
584.8	200.8	308.0	412.5	415.3
354.3	205.3	554.4	331.5	431.3
592.0*	241.3	373.3	402.2	205.7
373.7	472.4	302.4	400.7	200.6
331.4	245.0	354.0	358.0	282.1

* High and low values.

By definition

$$\log G = \frac{\sum f \log X}{N},$$

where G is the geometric mean, X is the value of the measurement, f its frequency, and N is the total frequency. This formula can be used if the frequency distribution of the logarithms is known. It is possible to get this distribution in a grouped form without getting the logarithms of the data.

Step #1. The Class Intervals. In order to avoid getting logarithms of each value of the data, we determine class intervals such that the ratios of the upper to the lower limit are constant. These are appropriate because then the corresponding grouping of the logarithms would have equal intervals. For the data in Table 1,

$$\begin{array}{lr} \text{(High value)} \quad \text{Log } 592.0 = & 2.7723 \\ \text{(Low value)} \quad \text{Log } 189.0 = & \underline{2.2765} \\ \text{Range of logarithms} = & .4958 \end{array}$$

Hence, in order to have seven classes, the range of the class interval of the logarithms should be about .071. Since the value .071 corresponds approximately to an antilogarithm of 1.18, 1.2 would be a

convenient ratio to use in determining the grouping (with unequal intervals) of the arithmetic values. Then the class intervals for the logarithms would have a range of log 1.2 = .0792. Thus, arbitrarily using 180 as the lower limit of the first class, we get the class intervals given in Table 2. These intervals are obtained by multiplying 180 times 1.2 to get 216, 216 by 1.2 to get 259, etc. (*A slide rule is adequate.*)

TABLE 2
THE FREQUENCY DISTRIBUTION

Class Intervals*	Frequency f	d	fd
180–216	1	−3	−3
216–259	6	−2	−12
259–311	12	−1	−12
311–373	14	0	0
373–448	11	1	11
448–537	3	2	6
537–645	3	3	9
Sums	50	0	−2

* Lower limits included in intervals.

Step # 2. The Frequency Distribution. The grouping of the data according to the unequal intervals in Table 2 is made in the usual way, and involves no more work than if the intervals were equal. It is important to see that this is also the grouping of the logarithms where the equal class intervals have a range of .0792.

Step #3. Calculation of the Geometric Mean. The so-called "short method" for calculating an arithmetic average from a frequency distribution can be applied in calculating the arithmetic average of the logarithms whose frequency distribution is given in Table 2. In this case the assumed mean is the mid-point of the class, log 311 − log 373, or 2.5323. Hence,

$$\log G = 2.5323 + .0792\left(-\frac{2}{50}\right) = 2.5291, \quad \text{or}$$

$$G = 338.2.$$

The above involves an application of the formulas

$$\overline{X} = \overline{X}' + i\bar{d}, \quad \bar{d} = \frac{\overline{X} - \overline{X}'}{i},$$

where $\overline{X}'$ is the value of the assumed mean, $\bar{d}$ the arithmetic average of the d's, and i is the range of the class interval.

THE DESIGN AND ANALYSIS OF METHODS FOR SAMPLING MICROCLIMATIC FACTORS

H. G. Wilm
U. S. Department of Agriculture, Forest Service[1]

The purpose of this paper is to outline the principles and use of a method for sampling factors—such as the amount of rain reaching the ground under a forest during a series of storms—which exhibit variations in both space and time. Solutions of two somewhat different sampling problems are discussed. One is designed simply to provide a sample estimate (average or sum) for any desired factor over a period of time. The other, with little loss of efficiency, provides a sample estimate for each of a series of short time units making up the period, as well as an estimate for the period as a whole. It is shown that these sampling methods are likely to supply a substantial gain in efficiency as compared to methods that have been commonly employed.

In recent years the intensive study of microclimate has added a new task to the science of climatology: that of sampling and comparing the average magnitudes of microclimatic factors on relatively small land areas. As illustrations, the climatologist may wish to compare the average amounts of evaporation and transpiration of water from fields containing different crops in a formally designed experiment, or the amounts of rainfall reaching the ground on watersheds or plots containing different kinds or densities of plant cover. In such studies the use of a single climatological station in each field or watershed fails to provide records adequate for reliable comparisons, and it becomes necessary to sample the desired factor or factors intensively in order to minimize errors associated with local variation.

By contrast, much of the earlier climatic research—like that in other fields—was designed to find out facts about the behavior of various climatic factors over a period of time; usually through the use of single climatological stations maintained at selected positions in space. With the aid of inductive reasoning, repeated experiments of this sort have contributed greatly to our fund of knowledge on the general behavior of these factors. When viewed in the light of our newer requirements, however, such studies are inadequate for deriving quantitative esti-

[1] Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado. Grateful acknowledgment is made to W. G. Cochran and F. X. Schumacher for their helpful suggestions and criticism of the methods described in this paper.

mates of the average magnitude of any factor on a land area large enough to contain local variations in microclimate.

The solution of this newer problem is not difficult; it is only necessary to sample each area by some standard technique, with continuous records obtained from a number of climatological instruments or stations installed at randomized points in the area so as to reduce sampling errors. This method may, however, be extravagant in cost, since it calls for a number of expensive instruments requiring considerable attention, and provides unnecessarily precise information on time variations while adequately minimizing the sampling errors which are associated with variations in space. Sometimes, in fact, the necessary number of instruments may incur prohibitive costs. One writer has even stated[2] that "it was early discovered that rain in the forest cannot be measured" because of excessive local variations in net precipitation under a forest canopy.

The purpose of this article is to outline a more efficient method for sampling microclimate and to discuss the statistical methods of analysis that are suited to it. With appropriate modifications, this method should be adaptable also to the sampling of other variables that are characterized by variations in both time and space. In brief, it requires the use of only a small number of instruments for each area to be sampled; these instruments are not placed at fixed positions in the area for the duration of the study, but are moved after each sampling observation to new randomized positions. Each individual observation contains sampling errors associated with both the time and the place at which it is taken. With properly designed sampling, however, unconfounded average or total values can be obtained for any desired time period long enough to permit thorough sampling in both time and space, and both types of sampling error can be simultaneously minimized with a good deal of efficiency. If these moving instruments are supplemented by one or more fixed stations—employed for the statistical control of variations associated only with time—, even short-time (as daily) average values may generally be calculated with satisfactory precision and little additional field work.

This method has been used in sampling net rainfall under a conifer forest[3,4] with a striking increase in efficiency as compared to the use of rain-gages kept at fixed locations. In a randomized-block experiment

[2] Rudolf Geiger, "Das Klima der bodennahen Luftschicht," *Die Wissenschaft*, vol. 78 (1927).

[3] C. H. Niederhof and H. G. Wilm, Effect of cutting mature lodgepole pine on rainfall interception, *Journal of Forestry*, vol. 41, pp. 57–61. 1943.

[4] H. G. Wilm, Determining net rainfall under a conifer forest, *Journal of Agricultural Research* vol. 67, no. 12, pp. 501–512, 1943.

a total of 80 gages, moved after each of 24 storms to sample new locations on each of 20 plots, provided data on average net and gross rainfall that were as precise as though over 400 fixed gages had been employed.

PRINCIPLES OF THE SAMPLING METHOD

Any factor to be sampled may be thought of as characterized by a true value μ for a given area and season or other relatively long period of time, and by variations within the area and period due to two general sources: a component η due to the time unit (as a day or storm) during which any single sample observation is taken, and a second component ϵ associated with the position of the sample observation in the area at that time. Then the magnitude of a single observation Y, taken during the ith time unit of the period and at the jth position in the area, may be expressed as:

$$Y_{ij} = \mu + \eta_i + \epsilon_{ij}. \qquad \text{I}$$

As stated in this equation, the magnitude of the time component is common to all observations taken in the area during any particular time unit, while the component of position is peculiar to each observation obtained during that time unit. If both are random and independent in the probability sense, the mean value of each or their joint value will tend toward zero with increasing numbers of observations.

In most investigations it is not practical to make an enumeration of all possible positions and time units in an area and period of time; instead it is necessary to take a randomized sample of these two populations. From such sampling the true average value μ and the true variance σ^2 of either population cannot be determined exactly; it is possible to obtain only a sample average and an estimate of the error of this average in approximating μ. If $Y_1, Y_2, \cdots Y_n$ are n values of a random variable drawn from a population whose true average and variance are μ and σ^2, then ($\bar{y}$ being the average of the observed values),[5]

$$\frac{S(Y - \bar{y})^2}{n - 1} = s^2, \text{ which provides an estimate of } \sigma^2. \qquad \text{II}$$

If the sample is made up of n observations, the variance of their average is equal to s^2/n; the variance of their sum is ns^2; and the square root of either value (standard error of the sample average or sum) provides an estimate of the failure of the sample statistic to equal the population value.

[5] For the symbolism employed in this paper, see G. W. Snedecor, *Statistical Methods*, 4th edition (pp. 34–35). Iowa State College Press, 1946.

In the sampling problem under discussion there are two categories of variation (time and space), each of which may be described by a variance. Then the variance of a single observation Y, randomized in both time and space, may be expressed as

$$s_Y^2 = s_\eta^2 + s_\iota^2 \qquad \text{III}$$

and the variance of the average of sample observations taken at k new places during each of n time units is

$$s_{\bar{y}}^2 = s_\eta^2/n + s_\iota^2/kn; \qquad \text{IV}$$

that is, the variance of the average is reduced in proportion to the number of observations taken in each category of variation.[6] If the sampling observations in time are taken repeatedly at the same set of k positions, then

$$s_{\bar{y}}^2 = s_\eta^2/n + s_\iota^2/k. \qquad \text{V}$$

And if a single instrument is left continuously at the same position (as in earlier studies), the variance s_η^2/n becomes zero, while s_ι^2 is not at all reduced. If the position variance is at all large, this procedure may cause a serious error in estimating the population value, simply because the available data describe only one point in the area sampled. It is evident that the sampling error of an average is likely to be minimized with greatest efficiency by taking observations at new places during each of a number of time units (equation IV).

Up to this point our discussion has dealt with random sampling in both time and space: a type of sampling which may be employed in the study of microclimate, although continuous observations in time may be more frequently desirable. In sampling microclimatic factors it is ordinarily necessary to visit each sampling point twice in order to obtain a single observation: once to install the instrument, and once to obtain the observation at the end of a short unit of time. For this reason and because values for short time units are frequently desired, it is usually most efficient to move each instrument to a new sampling location immediately after each reading is taken, so that the resulting observations follow each other without a break. By this means a complete enumeration is obtained of variations in time. Then, introducing the correction for a finite population, equation IV becomes

$$s_{\bar{y}}^2 = \frac{s_\eta^2}{n}\left\{\frac{N - n}{n}\right\} + \frac{s_\iota^2}{kn}\left\{\frac{K - kn}{K}\right\} \qquad \text{VI}$$

[6] Neglecting corrections for finite populations, which adjust the error variance of means to zero as n and k become equal to the whole population. (See F. X. Schumacher and R. A. Chapman, "Sampling Methods in Forestry and Range Management," *Duke University School of Forestry Bulletin* 7, 1942.)

where N and K are the total numbers of time units and positions in the two populations; and since $n=N$ when all time units are enumerated, the first right-hand term becomes zero.

DESIGN AND ANALYSIS OF THE SAMPLES

In the following sections are discussed the design and analysis of two variations of the general sampling method, as developed from the above principles. For the sake of clarity, the discussion includes only the simplest sampling designs; these may be varied and amplified to suit a variety of requirements. In the first sampling problem, the objective is to obtain a reliable average value for any desired factor on a land area over a period of time such as a month or growing season. In the second, the aim is to obtain relatively precise average values for the area on successive days or other short time units throughout the whole period as well as an average value for the period.

First problem: the estimation of an average value for a period of time.—If the investigator desires simply an average value for any factor, based on consecutive observations taken over a period of time such as a growing season, the only requirement is to obtain an unbiased estimate of this statistic and of its sampling error in space (as estimated by s_ϵ^2). These may be obtained by taking sample observations at two or more new positions, randomized over the area to be sampled, during each of a number of consecutive time units (as days) which add up to the whole period. Then the variations in time and space which have been sampled by this procedure may be partitioned into their components by a simple analysis of variance (Table 1).

TABLE 1

ANALYSIS OF VARIANCE IN THE SAMPLING OF MICROCLIMATE

Source of variation	*Degrees of freedom*	*Variances estimated by each mean square*
(1) Total	$kn-1$	
(2) Time units	$n-1$	$ks_\eta^2+s_\epsilon^2$
(3) Observations within time units	$n(k-1)$	s_ϵ^2

Since the sampling error variance of the average value obtained by this method contains only s_ϵ^2, for our present purpose it is necessary to calculate only the mean square for "observations within time units" in Table 1. From this mean square the standard error of the sample average may be calculated:

$$s_{\bar{y}} = \pm\sqrt{\frac{s_\epsilon^2}{kn}\left\{\frac{K-kn}{K}\right\}}. \qquad \text{VII}$$

This sampling method is illustrated by data obtained in an investigation of the quantity of net rainfall reaching the ground under a forest of young lodgepole pine (Table 2, "Net Rainfall": columns headed "Y"). In this study two standard rain-gages were placed at newly

TABLE 2

GROSS AND NET RAINFALL IN A FOREST OF YOUNG LODGEPOLE PINE

Storm No.	Rainfall—Inches			Storm No.	Rainfall—Inches		
	Gross	*Net*	*Calculated average*		*Gross*	*Net*	*Calculated average*
	(X)	(Y)	(E)		(X)	(Y)	(E)
1	0.00	0.08 .08	0.000	11	0.14	0.06 .12	0.100
2	1.23	1.01 .72	.074	12	.30	.28 .30	.228
3	.92	.76 .71	.725	13	.72	.63 .65	.605
4	.22	.20 .14	.164	14	.20	.11 .15	.148
5	.08	.05 .04	.052	15	.14	.06 .07	.100
6	.11	.09 .07	.076	16	.24	.08 .08	.180
7	.11	.04 .08	.076	17	.27	.28 .18	.204
8	.53	.38 .44	.413	18	.17	.16 .16	.124
9	.06	.03 .06	.036	19	.00	.06 .04	.000
10	.46	.26 .36	.356	20	.24	.13 .16	.180

Average total rainfall	6.32	4.82
Average rainfall per storm	0.316	0.241
Sum of squares "within storms" = 0.1107		
mean square = 0.0058		

randomized positions in the forest during each of 20 successive storms, so that at the end of the study period a total of 40 positions had been sampled. The observed sample value for the total net rainfall over the 20 storms was 4.82 inches of water; and the average net rainfall per storm was 0.241 inch. From equation VII, the standard error of this

average in estimating the true value for the area and period was calculated to be ±0.012 inch.[7]

The results of another study of net rainfall can also be cited to illustrate the precision of our short-cut method. In this experiment rainfall data were obtained at 20 positions in a pine forest, with the gages left at the same positions for a period of two years. Since the gage positions were not properly randomized, the resulting data may not provide unbiased estimates of sampling errors; but they can serve to show the benefits of more efficient sampling. For this purpose a consecutive series of 10 storms was selected, all occurring in a single month. Based on these 200 observations, the average net rainfall for these storms was 0.2973 inch; the smallest and largest storms averaged 0.009 and 1.036 inch; and the range from the smallest to the largest reading in a single storm was 0.58 inch. From these data 4 different samples of 20 readings were drawn, each composed of 2 readings taken at random without replacement from the 20 readings available for each of the 10 storms. The resulting sample averages were 0.3000, 0.3045, 0.3085, and 0.2950 inch. Thus even the poorest of these small samples provided an average value which deviated only 0.011 inch (3.8 per cent) from that provided by 10 times as many observations obtained with gages at fixed locations.

Second problem: the estimation of average values for each of a series of short-time units.—The relatively simple procedure required for solution of the first problem becomes somewhat more complex if a reliable average value is desired for each short-time unit in the whole period. Returning to equations I and III, we have seen that any single observation contains sampling errors due to the particular time unit and position at which it is taken; and that each of these components of error can be estimated by a variance. Hence it can be appreciated that, even though the average of a series of randomized observations may be quite precise (equations IV and VI), each separate observation or the average of only those obtained during a single time unit provides a relatively poor estimate of the desired factor for the whole area during that time unit. This fact is illustrated by the variation between the members of each pair of observations in Table 2, and by the large standard error of their average (pooled standard error = ±0.054 inch).

In order to obtain a more reliable average value for each time unit, we might take a considerable number of randomized observations during each of the series of units that make up the whole time period. But,

[7] Since variations in net rainfall tend to be correlated with the amount of rainfall in the storm, the logarithms of the data from this experiment might preferably have been employed in analysis. The natural numbers were used, however, in order to simplify the discussion of sampling principles.

as in the first problem, this procedure is unwieldy and may be made more effective. In order to increase efficiency we can proceed exactly as in the first problem, taking only two (or k) randomized observations during each time unit. For our new purpose, however—the statistical control of time variations—, one fixed station (or more than one, if desired) is employed in addition to the two moving instruments, with the new station maintained continuously at some convenient point in the area sampled. In order to minimize the chance of bias, this station should be carefully located and care should be taken to avoid the trampling of crops or any other influence that might be cumulative and peculiar to the fixed station.[8] Observations taken at the fixed station during successive time units will completely measure variations with time at this point; they will fail to do so for the whole area only to the extent that the true value for the whole area during each time unit interacts with the corresponding value observed at the fixed station. With the aid of regression analysis, these data will provide a basis for estimating the most probable average value for the space population during each of the time units which make up the period of study.

For this purpose a regression equation can be calculated to express the relation of observations taken at the moving stations (Y) to those taken at the fixed station. This regression should ordinarily be linear and quite strong; if it is linear it may assume the form

$$E = \bar{y} + b(X - \bar{x}) \qquad \text{VIII}$$

where E is an average value calculated for the space population during any single time unit; b is the linear regression coefficient computed from the corrected squares and products for "time units" (line 2 in Table 1); X is any single observation taken at the fixed station; and $\bar{x}$ is the general mean of all fixed-station observations.

From equation VIII can be computed an average value E for each time unit, corresponding to each value X observed at the fixed station. And the sampling variance of E may be calculated from an equation of the form

$$V_E = V_{y \cdot x}\left\{\frac{1}{kn} + \frac{(X - \bar{x})^2}{Sx^2}\right\} \qquad \text{IX}$$

where V_E is the desired error variance; $V_{y \cdot x}$ is the mean square for errors of estimate derived from the regression analysis (in line 2, Table 1); and Sx^2 is the corrected sum of squares for the fixed-station data. This

[8] Any such bias will not affect the average values obtained from the sample; but it might affect values calculated for individual time units, and it will be reflected in the magnitudes of the sampling errors for time-unit averages.

equation estimates the error of any value E in estimating the true average value for the population in space and time which is characterized by that single value for X from which the value E was calculated. As indicated by the right-hand term of the equation, the precision of the estimate E decreases as the deviation of X from the average value ($\bar{x}$) increases; where X is close to the average, E may be almost as precise as the average of all the moving-station data. When, on the other hand, X is very small or large, the calculated value E may be relatively unprecise. Hence this sampling method may not give satisfactory results where it is desired to obtain reliable estimates for extreme conditions.

It should be observed that, although the regression may have removed most of the effects of time variation on average values and their errors, the quantity $V_{y.x}$ may still contain variances associated with both time and space, as shown in line 2 of Table 1; but now the time variance estimates only a residual effect: the failure of the fixed station to isolate completely variations with time on the whole sampled area. Thus the mean square for errors of estimate may be expressed as

$$V_{y\cdot x} = ks_{\eta'}^2 + s_\epsilon^2 \qquad \text{X}$$

where $s_{\eta'}^2$ is the residual time variance. Incidentally, any bias associated with the fixed station should be expressed in the magnitude of this residual variance.

As it is developed above, equation IX is not entirely adequate for calculating the error of short-time values unless a single time unit and sampling position may be considered a minute portion of the populations in time and space which are being sampled. Ordinarily even the mean of a considerable number of sample observations in space may be so considered, as a single rain-gage or thermograph occupies only an infinitesimal part of even a single acre of land. In our sampling problem, however, the time population is completely enumerated; and even the residual error estimated by the variance $s_{\eta'}^2$ for a single time unit may represent an appreciable portion of the population of these time deviations. In the calculation of V_E from equation IX, therefore, it may be desirable to break down $V_{y.x}$ into its two components, and to adjust the time component for its sampling of a finite population. Then equation IX assumes the modified form

$$V_E = \left[ks_{\eta'}^2\left\{\frac{n-1}{n}\right\} + s_\epsilon^2\right]\left[\frac{1}{kn} + \frac{(X-\bar{x})^2}{Sx^2}\right]. \qquad \text{XI}$$

Obviously the use of this more complex error equation may induce

little gain in precision unless $s_{\bar{t}}^2$ forms a substantial part of the mean square for errors of estimate.

When the error variance of a seasonal average is calculated, all the right-hand terms of equation XI except s_t^2/kn become zero; as in our first problem, the average contains only sampling errors associated with variations in space.

This extension of the simpler method of sampling may again be illustrated by data obtained in a forest of young lodgepole pine (Table 2). From these figures on gross and net rainfall (X and Y) we calculated an analysis of covariance (Table 3), following the form set up in Table 1. Thence we obtained a linear regression equation which expressed the sampled relation of net to gross rainfall and from which, by inserting individual storm values for gross rainfall, we could compute the most probable values for net rainfall per storm on the whole study area. These calculated values (E), as shown in Table 2, are obviously more stable in their relation to gross rainfall than are the paired sample values for net rainfall, and provide a more reliable estimate of the true net rainfall per storm.

TABLE 3

NET RAINFALL IN A FOREST OF YOUNG LODGEPOLE PINE: ANALYSIS OF VARIANCE AND COVARIANCE

Source of variation	Degrees of freedom	Corrected squares and products			r*	Errors of estimate		
		x^2	xy	y^2		Degrees of freedom	Sum of squares	Mean square
(1) Total	30	3.7102	2.0705	2.0700		38	0.2818	0.0074
(2) Time units (storms)	10	3.7102	2.0705	2.5530	+0.067	18	0.1051	0.00021
(3) Observations within storms	20	0.0000	0.0000	0.1107	0.0001	20	0.1107	0.0058

Regression equation: $E = \bar{y} + b(X - \bar{x})$

$= 0.2410 + (0.801760)(X - 0.3100)$

$= 0.8018X - 0.0124$

* r = coefficient of correlation

† No regression available

‡ Not significant; $p > 0.05$

Using equation XI and the data in Table 3, we can estimate the sampling error variance of each of these calculated values:

$$V_E = \left[2s_{\bar{y}}^2\left\{\frac{20-1}{20}\right\} + s_t^2\right]\left[\frac{1}{40} + \frac{(X - 0.3100)^2}{3.7102}\right].$$

In this example

$$s_\xi^2 = 0.0058 \qquad 2s_\eta^2 = 0.0092 - 0.0058 = 0.0034.$$

Hence

$$V_E = \left[\frac{(0.0034)(19)}{20} + 0.0058\right]\left[\frac{1}{40} + \frac{(X - 0.3160)^2}{3.7162}\right]$$

which gives

$$V_E = 0.000226 + 0.002430(X - 0.3160)^2.$$

For the first storm measured, with a calculated average net rainfall of 0.060 inch and a gross rainfall of 0.090 inch, this equation becomes

$$V_E = 0.000226 + 0.002430(0.090 - 0.3160)^2 = 0.000350,$$

whence the standard error of the calculated average net rainfall for this storm is estimated to be

$$S_E = \pm\ 0.000350 = \pm\ 0.019 \text{ inch.}$$

Thus it can be seen that, by the use of only three instruments, we not only obtained a reliable average value for the whole study period with considerable efficiency, but also derived an average value for each short time unit in the period without much additional work and with relatively high precision. In sampling other variables or even net rainfall in other places, of course, the relative precision of this short-cut method will depend on the degree of correlation existing between observations taken at the fixed and moving stations.

SUMMARY

In the study of microclimate, the investigator may desire to obtain quantitative estimates of the average magnitude of some climatic factor such as rainfall or temperature on one or more areas of land. This objective has been attained in the past by establishing "stations" consisting of rain-gages or other instruments in each area and obtaining continuous records of the desired factor over a period of time. If enough instruments are installed in each area by some scheme of randomization, this method should give satisfactory results; but it may involve excessive costs for the necessary instruments and their care.

The cost of such sampling may frequently be reduced with little loss in precision by the use of a more efficient sampling method which is outlined in this paper. Instead of the considerable number of instruments which is ordinarily required for adequate sampling, this method

calls for only a small number of instruments—two or more—for each area to be sampled. If, for example, it is desired only to obtain an unbiased average value for the area over a period of time, two instruments may be employed. These are placed at randomized positions in the area, and moved after each of a series of short time units (as days or storms) making up the whole period. By this means variations associated with both position and time are sampled simultaneously, so that at the end of the period of study all time variations have been enumerated and a number of positions in the area have been sampled with a comparatively small amount of work.

While this sampling method provides relatively precise average values for each area and period, the resulting average values for the short time units are not so precise because each is based on the sampling of only two locations. If, however, one additional instrument (or more, if desired) is maintained continuously at a fixed position in each area for the whole study period, the records obtained from this instrument and the moving ones can together be employed for the calculation of more reliable average values for each short time unit.

With appropriate modifications, it should be possible to adapt these methods to the sampling of a variety of biological factors in addition to those associated with microclimate.

APPLICATION OF MACHINES TO DIFFERENCING OF TABLES

JACK LADERMAN, *Chemical Warfare Service, War Dept.*
AND MILTON ABRAMOWITZ, *Mathematical Tables Project, U. S. Bureau of Standards*

During the past twenty-five years, a variety of machines have been constructed for computational work required in business. Most of these machines, although designed primarily to meet the demands of business, can be applied to scientific computation. Mathematicians, statisticians, and other scientists engaged in computing are in most cases limited to the use of these machines. Although machines designed for specialized computational work exist, and no doubt a machine can be constructed for any type of mathematical computation, they are not available to most computers. Thus it is essential that existing business machines be applied to scientific computation. Great progress has been and will be made in the improvement of such machines if they are more universally utilized in scientific work and their inadequacies brought to the attention of the manufacturers.

The greatest application of business machines to scientific computation has occurred in the field of punched card equipment. The flexibility of the I B M punched card equipment has enabled those engaged in mass computational work to adapt this equipment to perform their tasks in a fraction of the time formerly required by other methods. While considerable progress has been achieved in this field, its full potentialities have not been realized. However, new applications of this equipment to scientific computation are being constantly discovered.

There are other business machines, although much more limited in operation in comparison with I B M equipment which, nevertheless, can be applied efficiently to certain computational work. One such machine is the Underwood-Elliott Fisher Sundstrand Accounting Machine, Model D, which can be applied to differencing of mathematical tables. The main purpose of this paper is to explain this particular application which is of great importance to mathematical computers. Differencing a table until a sufficiently high order difference vanishes or fluctuates within predetermined limits depending upon the order of difference, is the most efficient method by which professional computers check the accuracy of mathematical tables. However, there are many applications of a machine which can obtain differences of high order. For instance, such a machine can be used in the inverse operation of constructing a mathematical table from its differences, for comput-

ing moments by progressive summation, for interpolation, for subtabulation, and for numerical integration.

The first attempt to construct a machine to difference mathematical tables was made by Charles Babbage in 1823 who worked more than ten years and spent approximately $80,000 on his Difference-Engine without success. The main purpose of Babbage's machine was to compute tables from differences, but this operation is essentially the same as differencing tables. After a part of this machine was assembled in 1833, Babbage's thoughts became centered on a new idea, that of a more complicated machine which he called an Analytical-Engine. It is unfortunate that Babbage abandoned his Difference-Engine, for, as it turned out, he was unsuccessful in completing either of the two machines. Had he concentrated on the Difference-Engine, it is very likely he would have succeeded. Nevertheless, Babbage's attempt to build these two machines was not entirely wasted. There is no doubt that Babbage had a considerable influence on improving the precision of machine parts of his time and his work apparently influenced George and Edward Scheutz who succeeded in constructing during Babbage's lifetime a machine to calculate and print mathematical tables from differences.

Since Babbage's time many other attempts were made to construct a Difference-Engine but the machines were impracticable since they were too costly and too limited in their application to other computational work. Recently machines used in general computational work in business have been applied successfully to computing and differencing mathematical tables. The I B M punched card equipment has been used to difference tables[1] and to construct tables from differences.[2] However, only second order differences can be obtained in one run of the cards. To obtain higher order differences, the lower order differences are summary punched on a new set of cards which are run through the machine, their differences summary punched, etc. Another machine which has been used to difference tables is the National Accounting Machine.[3] This machine has been used by Comrie to obtain and print all the differences up to the 5th order. The Burroughs Accounting Machine has also been used as a differencing machine by Comrie in England and by Eckert in the United States. Ten-key adding machines

[1] Hartkemeier, H. P. and Miller, H. E., "Obtaining Differences from Punched Cards," *Jour. Amer. Stat. Assoc.* Vol. 37, p. 285-287, June 1942.

[2] McPherson, J. C., "Mathematical Operations with Punched Cards," *Jour. Amer. Stat. Assoc.*, Vol. 37, p. 275-281, June 1942.

[3] Comrie, L. J., "Inverse Interpolation and Scientific Applications of the National Accounting Machine," *Royal Stat. Soc. Supp.* Vol. III, Nov. 2, 1936.

have been used extensively during the past ten years by the Mathematical Tables Project to compute the nth differences of a table by adding $n+1$ successive values multiplied by the appropriate binomial coefficients. For example, the fifth difference is given by the relation:

$$-\Delta^5 f(x) = f(x) - 5f(x+h) + 10f(x+2h) \\ -10f(x+3h) + 5f(x+4h) - f(x+5h).$$

In the event that a computer does not possess an adding machine he will find that this differencing process can be performed with considerable success on any of the standard calculating machines such as Marchant, Friden, or Monroe.

Through the efforts of the United States Hydrographic Office, the mathematicians of the Mathematical Tables Project, U. S. Bureau of Standards, and of the Underwood-Elliott Fisher Co., the Model D machine has been set up to difference and construct mathematical tables. Since January 1945 these machines have been computing and printing all the differences up to the 8th order of 12 digit entries merely by setting the entries on a convenient 10-key keyboard and pressing actuating bars. The argument x, the entry $f(x)$, and the eight differences $\Delta f(x-1h)$, $\Delta^2 f(x-2h)$.... $\Delta^8 f(x-8h)$ are being printed on a single line across a 24″ paper roll at the rate of 120 lines per hour. The appearance of the results of differencing a table of x^5 is as follows:

x	$f(x)$	$\Delta f(x-1)$	$\Delta^2 f(x-2)$	$\Delta^3 f(x-3)$	$\Delta^4 f(x-4)$	$\Delta^5 f(x-5)$
6	7776	4651	2550	1230	480	120
7	16807	9031	4380	1830	600	120
8	32768	15961	6930	2550	720	120
9	59049	26281	10320	3390	840	120
10	100000	40951	14670	4350	960	120
11	161051	61051	20100	5430	1080	120
12	248832	87781	26730	6630	1200	120
13	371293	122461	34680	7950	1320	120
14	537824	166531	44070	9390	1440	120

The machine is a multi-register adding machine containing two add-subtract crossfooters and eight adding registers. The subtractions for differencing are performed in the crossfooters; each result is simultaneously printed and stored in one of the registers as well as being applied to obtain the next higher difference. The machine operates in accordance with a specially designed control plate set for this differencing operation. Other operations, such as constructing a table from its differences, can be performed on this machine by changing the control

plate. The designing and setting of a control plate has to be performed by the company but once a plate is set for a particular operation, it need never be changed except for normal mechanical repairs. Several different plates can be maintained with the machine so that the operations can be changed frequently without delay, the interchanging of two control plates being a simple process.

The application of the Underwood-Elliott Fisher Accounting machine, Model D, to the construction of mathematical tables is exceptionally efficient. If the function and all its differences are of the same sign and if the 7th or lower difference is constant, then the machine will compute and print automatically all the lower order differences and the function values without requiring the continuous presence of an operator. The operator would start the machine by setting up the first line containing the differences and the function value. The machine would then automatically repeat the constant difference, compute and print the lower order differences and function value, repeat the constant difference, etc. This is a considerable accomplishment. For example, one can evaluate automatically certain seventh degree polynomials over a long range of equally spaced values of the variable after setting eight quantities into the machine. The function values would then be computed and printed at a rate of three per minute. If the 6th or lower order difference is constant, the arguments can also be printed automatically. The appearance of the results in tabulating x^5 would be as follows:

x	$\Delta^5 f(x-5)$	$\Delta^4 f(x-4)$	$\Delta^3 f(x-3)$	$\Delta^2 f(x-2)$	$\Delta f(x-1)$	$f(x)$
6	120	480	1230	2550	4651	7776
7	120	600	1830	4380	9031	16807
8	120	720	2550	6930	15961	32768
9	120	840	3390	10320	26281	59049
10	120	960	4350	14670	40951	100000
11	120	1080	5430	20100	61051	161051
12	120	1200	6630	26730	87781	248832
13	120	1320	7950	34680	122461	371293
14	120	1440	9390	44070	166531	537824

The machine will operate entirely automatically for certain functions other than those having all their differences of the same sign, but in most cases when the differences change sign, or when the seventh order difference is not constant, an operator must be present to operate the machine. Function values can be computed and printed by inserting any column of known differences up to the seventh order and by press-

ing various actuating bars. The function values can be obtained at a rate of 120 per hour.

The foregoing Underwood-Elliott Fisher Accounting Machine, a standard business machine, accomplishes more than Babbage expected of his Difference-Engine and represents a marked improvement over other machines used to difference tables. The ease with which the operations of the machine can be changed, its comparative simplicity, easy manipulation, and low cost, greatly enhances its practicality and demonstrates how a machine of this type lends itself to specialized computational work. Unquestionably, many more machines of this general nature exist which appear to be unknown to scientific computers. There is no doubt that the great effort currently expended on computational work could be appreciably reduced by stimulating a wider knowledge of the capacities of existing business machines and by promoting a broader extension of their applications.

NOTE ON SAMPLING PROBABILITIES

HOLBROOK WORKING
Food Research Institute, Stanford University

H. G. ROMIG, reviewing in this *Journal*[1] a recent publication of mine, suggests that "It would have been wise for the author to have extended his remarks as many are confused by the fact that the probability of acceptance for product of p quality under a given sampling criterion differs from the probability of acceptance for a single lot of p quality. In fact, many have considered them identical."

The fact to which he alludes may be illustrated most simply by considering the question: what is the probability of finding no defective article in a sample of five drawn from a lot containing forty articles? The question may be answered on the basis of an assumption regarding either,

a) the quality of the lot, or

b) the quality of the universe from which the lot came.

If the assumption taken be that the *lot* has a fraction defective $p=0.1$, the probability in question is

$$P = \frac{36}{40}\cdot\frac{35}{39}\cdot\frac{34}{38}\cdot\frac{33}{37}\cdot\frac{32}{36} = .5729.$$

If the assumption be that the universe or *process* has a fraction defective $p=.10$, the specification of lot size has no bearing on the problem, and the solution is simply

$$P = 0.9^5 = .5905.$$

It is apparent that in this instance the probability corresponding to a specified process quality can be determined more easily than that corresponding to a specified lot quality. When the problem presented is less simple, as in the case of single sampling with an acceptance number $c>0$, or in problems of double or multiple sampling, the advantage of working from process quality is relatively greater.

Perhaps the ease with which assumptions regarding process quality can be justified on the ground of simplification has encouraged neglect of the question of principle. Another contributing circumstance has been that when such questions relate to industrial sampling, it is easy to point out that the sampler is usually concerned with a series of samples from an industrial process, and so apparently to justify considera-

[1] September, 1945, pp. 413–15.

tion of process average, ignoring the question whether the industrial process is truly a single process from the probability standpoint.

In any case, it seems worth while to emphasize that the best reason for founding probabilities of industrial sampling on process quality rather than on lot quality is neither a reason of expediency nor one arising from peculiarities of industrial sampling. Such reasons are worthy of notice, but they ought not to obscure the fact that consideration of process quality is to be preferred on grounds of principle. The principle, whether the question be one of the fluctuations of quality in industrial samples or of the sampling fluctuations of an average or of any other statistic or parameter, is that by founding the probability analysis on characteristics of the universe, serious difficulties of inverse probability are avoided.

FREDERICK L. HOFFMAN
1865–1946

FREDERICK L. HOFFMAN who died in San Diego, California, on February 23d at the age of 80 was almost the oldest among the veterans of the American Statistical Association. His earliest book printed as a monograph by the Association appeared fifty years ago and fifteen years later he was chosen for its President. For more than twenty years, 1896–1919, he was one of the most frequent and valued contributors to its publications.

Born and trained in Germany but coming to the United States before he was twenty he brought with him an industry, which more than Teutonic was almost titanic, an eager desire to improve man's estate especially by reducing its burden of sickness, poverty and crime, an unshakable conviction that if one was to do good rather than harm in such an effort one must study the successes and failures of the past and gather widely all relevant facts about the present situation, and that to this end one must rest heavily and constantly upon the statistical method. Becoming statistician of the Prudential Insurance Company of America about ten years after he landed and before he was thirty years old he quickly assembled in his office what was probably the largest collection in the country not only of Federal reports but also of the scattered and fragmentary state and municipal reports dealing with demography in the broadest sense of the term and began to pour out articles and books interpreting this material.

Upon the interpretation of what he gathered, for example whether the greater prevalence of tuberculosis among Negroes is due mainly to inheritance or to environment, or whether the increase in the deaths reported as due to cancer means that cancer mortality is really increasing, opinions might and did differ. His great service lay in focussing the attention of the public and of scholars upon many social evils and furnishing a basis of fact much needed for their successful analysis. This he did for mining and industrial accidents, and for many diseases especially prevalent in certain occupations or of importance because of the large number they enfeeble or kill. These studies were on tuberculosis, malaria and especially cancer.

Thirty-three years ago he read before a medical society a plea for a new organization to study and try to prevent cancer. As a result the American Cancer Society was formed almost at once with him as one of the founders and in his later years he gave most of his attention and devoted most of his publications to that field.

In the middle of the twenty year period which covered his contributions to the Association's Publications he read before it a paper on "Problems of Social Statistics and Social Research" which epitomized his life work before and after. He conceived social research as aiming at "the solution of the problem of poverty with all its resulting problems" and then outlined "a working plan of social research." This was arranged under ten heads: wage-earners' expenditures or budgets, periodical savings aiming to determine whether the prosperity of the masses is or is not increasing with a subdivision of this topic to determine how far the deposits in savings-banks came from wage-earners, safety of investment made by wage-earners, preventable industrial diseases, preventable industrial accidents, employments for persons physically impaired, morbidity and mortality of wage-earners, suicide, and finally "possible physical deterioration" of city populations and particularly of city children. To this paper he added eight short statistical appendices outlining some of the information he had garnered on these problems.

It would repay some student beginning work in these fields to review the subjects on which Hoffman pioneered in order to discover what progress later scholars have made. This would prepare one admirably to carry on his work.

Walter F. Willcox
Ithaca, New York

CARL SNYDER

An Appreciation

Carl Snyder, past president of the American Statistical Association, died on February 15, 1946, at Serena, Carpinteria, California, at the age of seventy-six after a long illness. He was born at Cedar Falls, Iowa, of an English mother and an American father. He attended the University of Iowa for two years, but his spirit was too restless to be confined to college routines and he gave himself in succeeding years an extensive and liberal education. At twenty he was the editor of a newspaper at Council Bluffs, Iowa.

He lived abroad in Switzerland and England for a number of years, writing on various subjects, but increasingly in the field of business and finance. From 1917 to 1919 he was a special writer on financial subjects for the New York Tribune, and in 1920 began fifteen years of distinguished service with the Federal Reserve Bank of New York, first as the organizer of the Bank's research department, and later as General Statistician. During this fruitful period he lectured frequently at Columbia University, the New School for Social Research, other Universities, and many economic gatherings.

He was president of the American Statistical Association in 1928 and was chairman of the Social Science and Economic Section of the American Association for the Advancement of Science in 1934. He wrote extensively for the Journal of the Statistical Association and other professional magazines and was the author of the following books which had a substantial circulation.

New Conceptions in Science	—1903
La Nuova Scienza	—1904
American Railways as Investments	—1907
The World Machine	—1907
Business Cycles and Business Measurements	—1927
Capitalism, the Creator	—1940

In any generation and in any field of knowledge the new ideas are the product of a few. Most people accept the fashionable ideas of their day, and their service (not a negligible one) is in distributing these ideas and proliferating them in greater detail. In economics a few years ago everybody wrote about the business cycle and spun out the theories of its causes and its consequences. At another period central banks were the fashion and we were told how our troubles could be cured just by having a central bank. To-day the fashion is compensatory

fiscal policy, which has its high priesthood and its ardent devotees. A few people, however, have the originality and the vigor to break out of traditional molds of thinking, and Carl Snyder was one of these. He was an enlightened non-conformist, and his mind cut through drab theory into the realities. He insisted on digging up more and more facts and in learning his economics directly from the facts more than from the textbooks.

The outstanding discovery that Carl Snyder made when he insistently examined all the old facts and persistently dug up masses of new ones was the impetus of growth in economic life under capitalism. He insisted on using the old-fashioned word "capitalism" instead of calling it "the enterprise system" or some other phrase that ducked the prejudices of politics.

He unearthed long series of figures, all of which consistently proved the law of growth. On ratio charts he plotted the figures with trend lines, fitted by eye or by mathematical formula, whichever did the job better. On the basis of these records of history he denied that the American economic system was decadent or doomed to a downward trend by over-saving or maturity.

In his search to see the whole picture and the real picture he joined to his studies of production studies of distribution and prices and wages and bank credit, trying to put together the composite in a way that made sense. Out of all this arose a new over-all measure of the fluctuations in the volume of trade and a new measure of the general price level as distinguished from wholesale prices. With these instruments there became possible for the first time a coherent quantitative measure of the equation of exchange.

Some of these series of measures were rough. The weights to be attached to different kinds of prices in the general price index or in different kinds of production and trade in the trade index had to be approximations. Dollar figures which were "deflated" to turn them into volume figures were subject to substantial error. So there were many critics who thought that these new discoveries were unprecise and departed from the pathway of exhaustive and critical scholarship, as indeed they did occasionally. But they came nearer to measuring the reality of the whole picture than very precise measures of bits and pieces.

It was from these studies that Carl Snyder developed a firm, even a passionate, belief that the Reserve System by its open market and discount policies could and should exercise a strong influence on business fluctuations. His findings and his convictions were no small influence on Federal Reserve policies under the unusual leadership of Governor

Benjamin Strong. The results showed something of both the truth and limitation in these principles. A new chapter was written in central banking theory and practice.

The best fruits of his thinking Carl Snyder put into his book, *Capitalism the Creator*. It is packed with the evidence he had accumulated. It is sprinkled with flashes of his curious, realistic, and vivid mind. It is withal loosely put together and is vulnerable to the critical barbs of the exact student, but it has in it more penetrating wisdom than tons of conventional economic writing.

One can do no better than conclude with a quotation from Virgil Jordan:

> "The burdens of prosperity and progress are borne by but a few creative spirits, insatiably seeking knowledge, recklessly risking danger, enduring endless labor to bring order and accomplishment out of the indifference of nature and the indolence of man."

W. RANDOLPH BURGESS

New York
April 1, 1946

BOOK REVIEWS

Edited by
OSCAR KRISEN BUROS
Rutgers University

Federal Dimensional Quality Control Primer: A Simplified Method for Applying Statistical Quality Control to Dimensions, Revised Edition. Providence 1, R. I.: Federal Products Corp. (1144 Eddy St.), 1946. Pp. 38. Paper. One to six copies, gratis; seven or more copies, \$0.25 per copy. *Two reviews follow:*

REVIEWED BY CASPER GOFFMAN
Industrial Mathematician, Westinghouse Electric Corporation
East Pittsburgh, Pennsylvania

THIS booklet gives an elementary presentation of the method of constructing control charts and of their use in controlling machining operations. Several excellent treatments of this subject already exist, notably the *A.S.T.M. Manual on Presentation of Data* and the American War Standards, Z1.1, Z1.2, and Z1.3. (Reviewed in this JOURNAL 40(231): 377–80 S '45.)

It is the reviewer's opinion that the present treatment would at first sight have greater appeal for the shop supervisor or inspector than most existing presentations. On the other hand, it offers no convincing explanation of the practical meaning of statistical control or of the relationship between control chart limits and specification limits. Indeed, the reader is likely to acquire a false understanding of the latter since the control chart limits on averages of 5 are likened to the 3σ limits of a frequency distribution for individuals. No mention is made of a frequency distribution for averages of 5. The reader is likely to believe—as many users of control charts do until shown otherwise—that a process will make good parts if it is in control and has control chart limits within the specification limits. The practical consequences of this misunderstanding can be drastic.

A short section (pp. 4–8) is devoted to a discussion of the variability inherent in a product and the use of the frequency curve to represent the degree of variability. Most of the remainder of the book is concerned with the control chart for averages and ranges.

Another technique, not a control chart technique, is considered on page 26. This consists of drawing action limits at calculated distances from the tolerance limits. The reviewer feels that this method merits greater attention than has been given to it in the book, especially since it has wide application to machining operations with tool wear.

Pages 32–34 consider a plan for operator participation in quality control.

REVIEW BY LLOYD A. KNOWLER
Associate Professor of Mathematics
State University of Iowa

ALTHOUGH *Dimensional Quality Control Primer* was probably written to aid in the sale of precision measuring instruments, it is worthy of rather careful attention by one interested in the field of quality control. As its name implies, the *Primer* is intended to furnish material suitable for reading by the uninitiated—"in no sense to be a complete treatise on the subject." References to some additional material are given at the end of the booklet.

Only the production or fabrication phase of quality control by statistical methods is discussed. This phase is further limited by a discussion of only the so-called average ($\overline{X}$) and range (R) charts. As might well be expected, but slight reference is made to design and specifications based on the process, and little or no reference to acceptance sampling, or to other types of control charts sometimes used in production or fabrication.

The following points, well known by workers in this area and often made by other writers, are probably worthy of special attention: (a) "Statistical Quality Control, in itself, is not a complete panacea for assuring the quality of a product. *Engineering and Management action, based upon the facts obtained from the statistical methods, is also necessary;*" (b) Statistical quality control is of value to many other industries such as textile, rubber, and optical where interest may not be focused on the mechanical parts; and (c) Statistical quality control is suitable for other than large organizations.

The booklet gives reasons for and states certain advantages secured from the use of dimensional quality control. Suggestions are given for the installation of statistical quality control with particular reference to selection of operation and dimension, the sampling procedure, the construction of $\overline{X}$ and R control charts and the interpretation of the results. An interesting example is used to illustrate these points.

In general the ideas of variation are rather well developed. However, figures 8 and 9 might be a little confusing since the frequency curve representing the entire lot in figure 8 is not the sum of the separate parts; and the 3σ limits and the average in figure 9 do not refer to the data set out on the chart. Again some disagreement might result from what was meant to be an elementary analysis of a car on a highway in figure 10.

A reader might question the wisdom of recording the data to four figures as on the work sheet in figure 11 when the deviation from the mean or a nominal would represent much less work, decrease the error hazards, and make calculations simpler. Also, the advisability of rewriting the largest value and the smallest value in computing the range is questioned. The numerical calculations on this work sheet might be rechecked: $\overline{R}=.0004$ is apparently obtained by dividing .0060 by 15. This discrepancy alters the numerical work and some of the analysis on the following pages.

As with many statistical papers, some new notation and variation from

somewhat standard procedures have been developed. For example, $\overline{\overline{X}}'$ and $\overline{R}'$ denote standard values instead of $\overline{X}'$ and R' as set by the American War Standards; crosses have been used for averages and circles for ranges instead of using the same notation for each chart; broken lines have been used as guide for averages, and vertical ones for ranges—this was defended but probably not too convincingly.

The difficult subject of control limits as related to tolerance or specification limits might very well be rewritten. No test formula is given (one might be implied, but it would be necessary to know theory or method to get it) to determine whether or not individual items meet specifications when the process is in control.

The examples and analyses in this booklet are much better chosen than in the first printing. The booklet is of convenient size, has an attractive appearance, and is easily read. The illustrations, diagrams, and pictures are interesting. The above comments are intended to be taken in a constructive manner. Many of the minor objections could be adjusted in the third printing.

It is recommended that the *Primer* be read by those interested in quality control by statistical methods. It represents a considerable amount of work.

Federal Products Corporation is to be commended for making more readily available, material on this important phase of quality control by statistical methods.

Methods in Climatology. *Victor Conrad* (Research Associate in Climatology, Harvard University). Cambridge, Mass.: Harvard University Press, 1944. Pp. xx, 228. $4.00. (London: Oxford University Press, 1945. 22s. 0d.)

REVIEW BY H. C. S. THOM
Senior Meteorologist, U. S. Weather Bureau
Statistical Laboratory, Iowa State College

THIS *is the first full-length book in English on methods in climatology.* It is devoted mainly to the application of statistical methods to meteorology, for climatology in a broad sense is simply statistical meteorology. The book definitely reflects the author's long experience as a meteorologist, and he is to be complimented for pioneering a text in this increasingly important field. The book covers a rather large but perhaps somewhat biased sample of the older climatological methods. Many who have been hoping for a text in applied climatology will be disappointed to find that this text does not fulfill their desires since it discusses no applied problems.

The book is divided into four main parts and follows a logical plan of development of the subject. Part I takes up general methods and is largely a discussion of elementary statistical methods. Meteorological or climatological elements together with their measurement and errors of observation are first discussed briefly. Next there are four chapters on statistical methods applied to these elements, the last two chapters of this series being devoted to

curve fitting and harmonic analysis. Part II covers the characterization of the various features of the meteorological elements. The statistical characteristics of the simple elements—temperature, pressure, and wind—are discussed in the first three chapters of this part making use of tools developed in Part I. Here the book departs slightly from its logical order for consideration of certain combined elements and then returns to complete the treatment of the simple elements, cloudiness and precipitation. Part III is concerned with the general problem of spatial comparisons of meteorological elements. After preliminary discussion of the tendency toward uniformity of records of adjacent stations, the problems of homogeneity of climatological series and the completion of such series are treated. A chapter on correlation and regression follows. This order does not seem entirely logical since correlation and regression are the natural tools for the analysis of comparisons and might better have been introduced at the beginning of this part. Graphical methods of comparison, related problems in anomalies, and the special problems of wind occupy the next three chapters. Part III is completed with very brief chapters on air mass climatology and climatic indices. Part IV discusses the climatography which is the general description of the climate by various modes of representation. There are appendices devoted to the forms of climatic tables, tables required in the text, and a handy table of day numbers.

The main defect of the book is the inadequacy of the discussion of statistical methods. While the author mentions modern statistical methods largely by references, few of these methods have found their way into the text. The distinction between the sample and the population is not made clear, and tests of significance are barely mentioned. The fact that statistics themselves are random variables is not explained and little use is made of standard errors of statistics. The statistician will readily recognize the statistical method as that of twenty or more years ago even though references are made to more recent papers on climatological method.

Specific criticisms of the statistical method are principally those of vagueness, inadequacy, or omission. On page 19 the problem of choosing the size of the class intervals is summarized with "The choice of the size of class intervals is dependent upon the number of observations included in the sample and on the degree of exactness required." This gives the reader no method for making the choice. Again on pages 24–29 median quartiles, and deciles are discussed separately but could better have been treated generally as quantiles which are important in climatology. No mention is made of the standard errors of these statistics and their variability is described by "these methods assume their real value only if the size of the sample is so large that half the number of its items represents a series of many hundreds." On page 34 the "more correct" description of the normal distribution of the footnote might have been incorporated in the text. At the bottom of page 35 *et seq.* the fact that distributions must be normal in order that such an interpretation of the standard deviation may be made is too vaguely ex-

pressed. Many distributions in climatology are not normal and care must be exercised in the use of the standard deviation. Neither tests of goodness-of-fit nor errors of forecasts are discussed in the chapters on curve fitting. The discussion of curve fitting and regression could have been simplified by use of the analysis of variance. An early introduction of the analysis of variance would also have meant a great improvement in all of the statistical discussion. In the chapter on comparison of observational series clearer and more meaningful exposition could have been attained by extensive use of regression theory. On page 152 the non-normality of the distribution of r is not mentioned and the student who has acquired the level of statistical knowledge offered by the rest of the book might get into trouble using the probable error of r. A number of statistical concepts important in climatology such as contingency tables, autocorrelation, the tendency to normality of the distribution of certain statistics, special distributions of climatology, the chi-square test and others, and goodness-of-fit of frequency curves are either not mentioned or inadequately treated.

On the climatological side, the book is definitely continental in scope. References to papers of a theoretical nature are largely to the author's own papers or to those of continental writers. The excellent modern works of the Russians have been ignored. To be sure language difficulties here are great; however, abstracts and translations have been provided for a number of the important Russian contributions. References to American and English theoretical works are also few. Both of these countries have been more active in the development of modern statistical methods than the continental countries and both have consequently applied these more extensively to meteorology. Such concepts as the recurrence interval, Thiessen method of weighting, area-depth curves, and others all typically American and important methods in climatology are omitted. The decision for inclusion or omission, of course, remains the prerogative of the author, but this text only runs to 215 pages, a short book for such a large subject.

The author characterizes his own book well in his conclusion which the reader would do well to read first:

> The reader who runs over the pages of this book is perhaps disappointed in not finding this or that special method. In the first pages, it was shown that the number of derived elements is not limited theoretically; thus completeness is unattainable.
>
> Moreover, it is intended to present only a system of methods which facilitates the step from the qualitative to the quantitative. Many ways are offered for attaining this purpose, and they all should lead not only to an exact and comparable description of the climate but, in the end, to a physical explanation.

The number of elements being unlimited does not mean that the number of methods cannot approach adequate completeness. The second paragraph of the conclusion is a broad claim hardly justified by the contents of the book.

This book is finding and will continue to find considerable use as a source book for the qualified statistician working in meteorology and is highly recommended for this purpose. It is not recommended as a text or handbook of climatological methods.

Cleveland Market Data Handbook, 1945 Edition. *Howard Whipple Green* (Director, Real Property Inventory of Metropolitan Cleveland). Cleveland 15, Ohio: Real Property Inventory of Metropolitan Cleveland (1001 Huron Road), 1945. Pp. 69. $2.50.

REVIEWED BY FRANK STROHKARCK
Research Department, The Curtis Publishing Company
Philadelphia, Pennsylvania

THIS handbook, an annual publication, is well known in its field. As the title suggests, the handbook is prepared for the use of marketing men interested in the Cleveland area.

It is a compilation of statistical data drawn from federal and local sources and presents valuable statistics derived from local sources heretofore not published.

The 69 pages which comprise the handbook are made up of 19 pages of promotional text, 34 pages of charts, 13 pages of tables, and 3 pages of index.

The author, as in past publications, has attempted to show too much information in some of the charts; e.g., the bars of the charts are often divided into four or more components. A few of the curve charts are also confusing because of the detail of data shown.

The handbook will be very useful to those interested in obtaining social and economic statistics concerning the Cleveland area.

Statistische Methoden für Naturwissenschafter, Mediziner, und Ingenieure. [Statistical Methods for Natural Scientists, Medical Research Workers, and Engineers.] *Arthur Linder* (Professor of Applied Mathematical Statistics, University of Geneva; Instructor at the Cantonal Technical University in Zurich). Basel, Switzerland: E. Birkhäuser & Co., 1945. Pp. 150. Cloth, Frs. 18.50; paper, Frs. 15.50.

REVIEWED BY HENRY SCHEFFÉ
Associate Professor of Engineering
University of California at Los Angeles

CONSIDERING the general abundance of scientific books written in German, one is puzzled by the lack heretofore of any expounding the familiar (to us) statistical methods employing the t, F, and χ^2 tables. Since this book—an elementary exposition in German—will be of use to few readers of this journal, we shall omit a detailed summary and a listing of errors,* and instead remark on some aspects which may be of wider interest.

* There are a moderate number.

One of these is the organization of the book, which turns out to be quite successful: It is divided into two parts, the first of which, requiring little mathematical knowledge, is in spirit and content similar to R. A. Fisher's *Statistical Methods for Research Workers*, but covers less ground, while the second offers mathematical derivations, mostly by Fisher's geometrical methods, of all the distribution theory forming the basis of the first part.

In spite of its title there is little in this book aimed directly at engineers. Treatment of acceptance sampling is lacking, that of control charts and curve fitting is highly insufficient. The one example of curve fitting in this book is apt to give the engineer or physicist a poor impression of what statistics has to offer him here: A *linear* graph is fitted for the regression of y on x where x is the velocity at which the brakes are applied in an auto and y is the distance required to stop. The range of data is considerable, x varying from 4 to 25 mi/hr, the variance of y is thus far from constant, yet equal weights are used. Later it is decided by a statistical test that the straight line is satisfactory as a regression curve. The reader with a little knowledge of mechanics will be left skeptical or confused. This might have been used as an example of the powerful procedure in which the form (or plausible forms) of the regression curve is determined from physical considerations, after which the parameters (or decision between forms) are determined by statistical methods. The advisability of considering a transformation to logarithms in physical problems could also have been illustrated here. The moral is that it is dangerous to expound applications in a subject matter field without a good feel for its elements.

An interesting difference in mathematical education may be inferred from the fact that while for the first part, according to the introduction, "a knowledge of the simplest algebraic formulas suffices," single and double subscript and summation notation is freely used there. This would trouble most American students who have had a year of calculus. One also finds it strange that in a language so adaptable to the coinage of new words the same word (Streuung) is used for variance and standard deviation. Actually it is used here with five different meanings: for population and sample values of these two quantities (although the distinction is preserved in the mathematical symbols used) and also for the more general idea of dispersion, as when we say "measures of dispersion" (Streuungsmasse).

Operating characteristics (power functions) of tests, and estimation by confidence intervals are not mentioned, although the author comes very close to the latter in the case of the parameters of the binomial and Poisson distributions, and all the necessary distribution theory for many examples of confidence intervals is at hand in the book. Indeed, except for brief mention of a control chart for the sample values of standard deviations, there is nothing that could not have been written twenty years ago after the appearance of Fisher's famous book. Nevertheless, this is a very attractive little book and its author has done a great service to science: He merits the gratitude not only of all experimental workers who read English with less ease than

German, but of statisticians everywhere who should be happy that these methods have been made more widely available, and may even hope that some who would not otherwise have been attracted may be stimulated to further study and perhaps eventual contributions to statistics.

Time Series Analysis: Smoothing by Stages. *Lewis A. Maverick.* San Antonio, Texas: Paul Anderson Co., 1945. Pp. x, 153. $3.00.

REVIEW BY LEONID HURWICZ
Research Associate, Cowles Commission for Research in Economics
The University of Chicago

THE main objective of this book is to present the mechanics of a method of handling time series. This is accomplished by a general description of the procedures to be used, followed by a number of illustrations taken from the field of economics (San Francisco real estate activity; wheat prices and production; crude petroleum output; production of cotton, pig iron, and Portland cement; Erie Canal freight).

A partial justification of the method is contained in the preface along with some references to the author's earlier contributions.

Consider a time sequence of observations X_t on an economic variable X at time t. The time unit may be a year or say, a quarter of a year. Then, it is (implicitly) assumed by the author that the observed values may be decomposed into (1) the trend (possible including "long waves"), (2) the major cycle, (3) the short cycle, (4) the annual cycle (the seasonal), and (5) the irregular (random?) remainder. Write this as

$$X_t = \sum_{j=1}^{5} F_j(t)$$

where F_j is the j-th component in the above listing. (The alternative formulation in terms of products rather than sums gives rise to analogous procedures.)

Defining

$$G_i(t) = \sum_{j=1}^{i} F_j(t),$$

the method of "smoothing by stages" may be described as that of fitting $G_3(t)$ to X_t, $G_2(t)$ to $G_3(t)$, etc. The G's are called "smoothing lines." $G_3(t)$ ("smoothing line A") eliminates the seasonal fluctuations and the irregular remainder, but retains the two cycles and the trend; $G_2(t)$ ("smoothing line B") is free of the short cycle, but still retains the major cycle and the trend; $G_1(t)$ ("smoothing line M") is supposed to contain the trend only.

Fitting is done by moving averages or freehand. A good deal of space is devoted to advice on the methods of fitting. Once the smoothing lines have been drawn, it is possible to obtain the cyclical and seasonal components by subtraction (or, in the multiplicative case, by division).

The properties of the separate cyclical components can now be studied. Further, a claim is made that such decomposition will help correlation studies—either for various components of a given series or for the same component of several series. This is designed to help find certain causal economic relationships. Finally, extrapolation of the smoothing lines is suggested for forecasting purposes. The author is careful to warn with regard to both correlation and forecasting that mechanical procedures must be supplemented by economic theory and corrected for anticipated departures from past patterns.

The author does not use any of the tools of modern statistical inference. There is no explicit statement of the properties of the universe or precisely what it is that is being estimated or tested. Sampling problems are ignored.

Apparently in ignorance of the work recently done in the field of time series (von Neumann, Koopmans, Haavelmo, Mann, Wald, and many others), the author repeats, without taking sides, the once widespread fallacy of nonapplicability of correlation methods to time series.

There was a time when the economist's distrust of statistical procedures was amply justified. But the well-founded objections which had been raised against mechanical application of least squares regressions to economic data have stimulated the development of new methods which utilize economic theory and follow the principles of statistical inference (cf., e.g., Tjalling Koopmans, "Statistical Estimation of Simultaneous Economic Relations," this JOURNAL, December 1945 and references cited therein).

It may be that some elements of the author's method will turn out to be acceptable, if not optimal, for certain purposes. But a proper appraisal must await rigorous formulation of the model to which the method is to be applied.

An Introduction to Educational Statistics. *C. W. Odell* (Associate Professor of Education, University of Illinois). New York 11: Prentice-Hall, Inc. (70 Fifth Ave.), 1946. Pp. xiii, 260. $3.50. *Two reviews follow:*

REVIEW BY PAUL J. BLOMMERS
Director of University Examinations Service
University Examiner and Registrar
State University of Iowa

PROFESSOR Odell's stated purpose in writing *An Introduction to Educational Statistics* was to provide a textbook suitable for use in a one semester course designed to develop in students of education the ability: (*a*) to select appropriate procedures and measures, (*b*) to compute measures, (*c*) to interpret measures reported by others, (*d*) to associate terms and concepts, and (*e*) to present results.

The content, except in the matter of emphasis, does not differ markedly from that of other statistics textbooks written for students of education and

psychology. Briefly stated, it deals with the common descriptive statistics, with measures of relationship, and with sampling error theory. The relative amount of attention given these rather broad and overlapping divisions corresponds to the order in which they have been named. The treatment is (as it must be in a book of this type) nonmathematical.

The difficulties experienced by those who have undertaken the task of preparing nonmathematical textbooks on statistics are generally recognized and appreciated. Wordiness is almost inevitable, and rare indeed is the book of this type which has escaped without some involved and awkward sentences—not to mention involved and awkward expositions. Professor Odell's book is somewhat less mathematical than most volumes of this type and consequently it is not surprising to find such sentences and expositions occurring in it. To illustrate the point, a few examples are submitted: On page 50 is an explanation of the procedure for computing the arithmetic mean of grouped data by the so-called "short method." The description of the entries in the d column is concluded as follows: "In other words, they show how many intervals above or below, that is, larger or smaller than, the assumed mean class each of the other classes is." On page 127 is the sentence: "In general, if a subtest is uncorrelated with the other subtests composing the whole test, the coefficient of correlation between scores on it and on the whole is $\sqrt{1/n}$, in which n is the number of times as long the whole is as the subtest." And on page 131 in a section on interpreting coefficients of correlation: "by dividing $\sqrt{2-2r}$ into values of the measure of deviation employed .5 less and .5 more than the difference in position concerned, using the results as entries in the first column of such a table of the normal curve as is given in appendix, finding the corresponding areas between pairs of ordinates at the given deviation distances, and finally doubling them, the per cents of cases falling within divisions the positions of which differ by the given deviation distances may be found. *This sounds more difficult than it is.*" (Italics mine.)

Perhaps these sample sentences (it is not possible to cite here examples of lengthier expositions) enhance the contention of Professor Helen M. Walker that it is impracticable to present statistics without resorting to some elementary mathematical concepts and that individuals desirous of studying statistics and lacking a knowledge of these concepts had best take time to learn them first. Or perhaps these sample sentences suggest careless editing. They are not to be construed as typical of the writing of the entire volume. The fact remains, nevertheless, that they occurred with sufficient frequency to cause irritation.

Sampling error theory is discussed in the book's final chapter. The treatment is brief. This is surprising in view of the importance of the theory in educational statistics and in view of the liberal amount of space devoted to topics much less complex.

In postponing consideration of sampling theory to the final chapter, the author follows a pattern set in several other recent books of this type. It

seems unfortunate that at least a portion of this material was not introduced earlier in the volume so that advantage could have been taken of every possible opportunity for applying the theory in a variety of situations. It is also invariably true, when sampling theory is thus postponed, that the *student must cope with problems in which sampling is implicit without* benefit of a preliminary discussion of the theory. In Professor Odell's book, such problems range from the extremes of smoothing frequency distributions (page 38) to the chi-square test for goodness of fit (page 210). In the latter instance the use of chi-square as a test of the hypothesis that a given set of measures is a random sample from a normally distributed universe is described without benefit of any preliminary discussion of sampling theory. In this connection it should be observed that Professor Odell chooses to ignore the work of R. A. Fisher, and to use the number of classes less one as the number of degrees of freedom despite the fact that both the sample mean and standard deviation were used to obtain the theoretical frequencies. Professor Odell is somewhat inconsistent in this position, however, for later in discussing contingency tables (still without benefit of an introduction to sampling) he uses $(r-1)(c-1)$ as the number of degrees of freedom for chi-square.

When the discussion of sampling theory is finally reached, one learns (p. 250) that the author does not approve "the use of degrees or levels of confidence in stating the significance of results." Instead he prefers to state the number of chances to one that the obtained statistic is within a specified distance of the corresponding hypothetical value. In making such statements, several instances occurred in which the "chances to one" phraseology was improperly applied. To illustrate, on page 244, the discussion of an example is concluded with the statement that the chances are 3.7 to 1 "that there is a real difference in the direction indicated." In this connection it should be observed that Professor Odell advocates the use of the area under only one tail of the distribution in stating "chances." Again his position is inconsistent, however, for in the case of the t test applied to means, he employs the area under both tails.

Attention should be called to the fact that in Professor Odell's opinion most of the measures used in education are reported to the next smaller unit rather than to the nearest unit. Throughout the volume, the limits of the classes of grouped distributions are defined so as to be consistent with this view. It should be noted, however, that this view cannot be reconciled with present day theory of educational measurement.

Some prospective users of this text may find to their liking the author's attempt to "enrich" the students' statistical vocabulary. On page 83, for example, he introduces the need for a measure of *variability*, "also called *variation*, *dispersion*, *deviations*, *departure*, *discrepancy*, *spread*, *scatter*, and *fluctuation*." In the opinion of the reviewer the value of this practice is questionable.

It is impossible within the space limitations of this review to give a com-

plete criticism of this book. A few final objectionable features will be cited briefly: (*a*) The author departs from accepted notational practice and uses to represent standard error (Chap. 14). (*b*) The correlation ratio is said to "summarize the deviations of the data from the curved line that *best fits* [italics mine] the means of the rows or columns . . . " (p. 179). (*c*) The median deviation is said to share many of the merits and weaknesses of the standard deviation because it "is regularly found by multiplying the standard deviation by a constant" (p. 100). (*d*) A scheme for changing ordinary ranks to the "same basis" (i.e., to percentile ranks) is presented as the sole requirement for making ranks comparable (p. 77). (*e*) The averaging of percentile ranks is presented as leading to the "average rank." That this "average rank" may not correspond to the percentile rank of this average is not mentioned (p. 78). (*f*) A procedure for correcting the standard error of the mean for the effect of errors of measurement was presented, and the corrected value used in a test in which the fact that errors of measurement were also involved in the obtained mean was ignored (pp. 238 ff.). (*g*) It is suggested that bias in sampling may be revealed by drawing, "in the same manner, but independently, several samples of equal size," and comparing them (p. 230).

REVIEW BY P. E. VERNON
Lecturer in Psychology, University of Glasgow

FROM many points of view this is an admirable book. The reviewer cannot call to mind any other which contains so complete a collection of statistical formulae, with precise directions as to their applicability. The conciseness and clarity of the definitions are a joy to the reader. As a textbook for beginners, however, there is more room for doubt. Professor Odell explicitly puts it forward as covering a one-semester course, and claims to emphasize "practical computation, use, and understanding, rather than derivations and theoretical considerations." No higher mathematics is required.

Possibly the teachers and students whom Professor Odell has in mind are more highly selected than the reviewer's psychology and education students, the great majority of whom would be confused and put off by the very abstractness, precision and comprehensiveness of his treatment of the subject. Standard books such as Garrett's and Lindquist's are perhaps more verbose and in many respects cover less ground, but they would appear to provide the beginner with a better understanding of the uses of statistical methods. In the reviewer's opinion, therefore, this book should chiefly be employed by the research worker for reference, when he has already gained some general comprehension of statistics from other sources. Professor Odell does, of course, restrict his field to some extent by omitting, for example, analysis of variance and covariance and factor analysis, and by giving only very brief accounts of multiple correlation, small sample formulae, principles of experimental design, test reliability, item analysis, and other topics out-

side the usual run. But surely it is more important for the "consumer" to possess some understanding of just these topics than to know, say, the formulae for the standard errors of thirty statistical measures, or several ways of correcting r, S.E.$_m$, etc., which very possibly should be but seldom are used in everyday practice.

A few more detailed points may be picked out for comment. Although chi-squared is one of the most important, and most widely applicable weapons in the statistician's armoury, it is only introduced, in a limited way, towards the end of the book. Actually the coefficient of mean square contingency is described in an earlier chapter without resorting to chi-squared—a coefficient which has little if any practical value, since product-moment r based on mean class values is greatly preferable. The simplest method for the beginner to calculate product moments from grouped data is probably Godfrey Thomson's, which involves the variance along the diagonal of the correlation table; this is not mentioned. The stock method for biserial r, using p, q and z, would seem considerably easier than the method Odell provides. The quadratic formula for tetrachoric r is given without sufficient warning as to the inaccuracy of its approximations. Cyril Burt's trigonometrical approximation is probably closer, and certainly quicker when graphs are not available.

It would be unfair however to list these criticisms without acknowledging how frequently Professor Odell provides novel and extremely useful variations of standard formulae. The reviewer was interested also in his advocacy of the 10th–90th percentile range as one of the best indices of dispersion, for it turned out to be a particularly valuable measure in vocational work during the war in the British Forces. Several of the statistical tables, again, are rearranged in convenient forms, though they tend on the whole to be too coarse for the practitioner; thus the normal curve table only gives heights, areas, etc. for abscissae from 0.0 to 5.0 by intervals of 0.1. There are good exercises and problems at the end of each chapter, but no answers. The typography is extremely clear, and only two misprints have been noted, namely the omission of cross-lines through sets of 5 tallies on page 24, and the absence of σ in the biserial r formula on page 185.

Fundamentals of the Theory of Statistics: Vol. 1, Elementary Statistics and Applications; Vol. 2, Sampling Statistics and Applications. *James G. Smith* (Professor of Economics) and *Acheson J. Duncan* (Assistant Professor of Economics). (Princeton University.) New York: McGraw-Hill Book Co., Inc., 1944, 1945. Pp. x, 720; xii, 498. $4.00 per volume.

REVIEW BY JOHN WISHART
Reader in Statistics, University of Cambridge

IT HAS been a matter of considerable interest to the reviewer to examine a modern text on statistics as his first academic task after six years' absence on war work. The first thing to note is that these books do not simply

represent volumes 1 and 2 of a connected treatise. They are intended to be self-contained expositions, the first elementary and the second advanced. The books have already been reviewed by W. Edwards Deming (this JOURNAL 40(232): 534–6 D '45) largely from the point of view of the need for an exposition of sampling technique. We are here, however, up against two definitions of the word "sampling." The book does not profess, in this reviewer's judgment, to go very deeply into the technique of sampling methods and surveys, and it may therefore be useful for readers to have described to them the achievements of the books as a combined theoretical text, especially as what is meant by "Sampling Statistics," as opposed to "Elementary Statistics," is really tests of significance, exact so far as they can be made exact, as worked out by mathematicians for finite or "small" samples, compared with the more classical "large sample" theory.

The first book is arranged on more or less orthodox "nonmathematical" lines, the inverted commas being used to express the fact that while there is a great deal of algebra (and also geometrical illustration) in the working out, for example, of partial regression coefficients, and some deep mathematical and logical concepts involved in the study of probability and its application to the confidence interval, the book is concerned with calculating the descriptive "statistics" rather than testing their significance. A very great deal of groundwork is covered, and the book is particularly full on the questions of the gathering and presentation of statistics. The reader coming fresh to the subject will learn a great deal of what statistics is about, although it is quite clear that in these days this will not be enough, and he will therefore be required to go on to the subject matter of the second book. This is implicitly recognized by the numerous forward references to the second book. This raises the question of whether it is intellectually satisfactory to leave so many questions unresolved until a more advanced stage in tuition is reached. To some extent the books fall between the two stools of providing self-contained texts for an elementary and a more advanced set of readers. One illustration of this is that the study of the simpler "statistics," i.e., mean, standard deviation, etc., is followed up by tests of significance and the use of the confidence interval, by means of the "large sample" standard errors of the statistics, whereas from chapter 13 onward there is no more mention of such tests. Even the elementary student should be told how to test the significance of a simple regression or correlation coefficient, by means of its standard error or in other, more up-to-date, ways. There is a formula for the standard error of z (where $r = \tanh z$) on page 320, but it has not been applied in the next chapter.

It might have been better to have planned the whole as one standard text, while indicating at the same time what sections could be omitted on a first reading, or by those who were prepared to take on trust the derivations of mathematical formulae, being content if they know the basic principles on which tests were based, and how they worked. Living, as the reviewer does, in a country where authors are still rather rigidly controlled in the amount of

paper they can use, it is apparent that such a method would have avoided a good deal of the repetition that has gone into the second book with the commendable object of making it self-contained. The second book does not appear on a first reading to be nearly as good in construction as the first. True, it contains a great deal of sound argument, not usual in the ordinary text, on the inferences that can legitimately be made concerning the population parameters from a knowledge of the sample statistics, and it is clear that the authors have studied this question and the work of Pearson and Neyman very thoroughly. But the book is somewhat uneven as a whole, conveying the impression that questions which have specially interested the authors, and which have doubtless been in consequence worked out in great detail in the laboratory, are reproduced in full, while in others the treatment is somewhat sketchy. Since there is little in the modern theory which is not dealt with, it is surprising to find such a small author index, and to note the absence of a number of eminent names, surely the easiest way of finding what has been left out of the book.

It must have surprised the authors to read in W. Edwards Deming's review that "the second is not a textbook in mathematical statistics." The present reviewer thought on examining it that it was intended to be such a text; the present criticism rather refers to the uneven "weights" attached to the different aspects of the subject. It is doubtful whether a long explanation of the distributions of samples of two by means of an arithmetical model is needed when the full composite distribution of mean and variance is later derived; on the other hand it is somewhat startling to find in chapter 12 that mathematical methods are laid aside and Fisher's distribution of r merely quoted. Somehow one feels that a text should be either the one thing or the other. *Either* these distributions should be quoted, explained and illustrated, giving the same treatment to all, *or* they should all be derived as well. Another example is the goodness of fit distribution. The impression is left on the reviewer's mind that the authors are not yet sufficiently at home in the synthesis, and therefore exposition, of the formidable amount of mathematical theory that goes to make up the present-day theory of statistics. There is no reason, however, why with a good deal of revision in a subsequent edition the book should not prove useful for the class of student for whom it is intended. The opportunity could then be taken to correct a fairly large number of minor errors of one sort or another that have no doubt already been spotted. A few of the more important may be set down here. On pages 125–26 the .95 point has been read from table 8 instead of the .98 point; page 142 (footnote) suggests that 11 degrees of freedom, and not 8, have been taken; table 24 has not been calculated very accurately, and there is at least one major error (.2033 for $p_1 = .105$ should be .2136), also the mean should be p_1 in the footnote to the table; the formula in the heading of table 41 is wrong, while it is doubtful whether table 42 should be used for N as low as 11 except as a rough test—Neyman and Pearson suggested it for N greater than 50. On pages 410–11 the authors are evidently referring to a different table

from table 9, since in the latter the values for $n_1=0$ appear, which makes interpolation rather pointless, and incidentally shows up the weakness of the interpolation method advocated. One wonders why Fisher's harmonic method on z, half the natural logarithm of F, cannot be used. The authors have incidentally used the .01 values throughout instead of the .05 values in this illustration.

Lastly, a word about analysis of variance. The authors began promisingly in the first book by illustrating analysis of variance in the simple problems in terms of population parameters. It was therefore somewhat disappointing to find the method dealt with very sketchily in the second book, and treated as a somewhat limited tool for special tests. It is not known why they refer to "a so-called 'analysis of variance'." The name is well established and is used by the authors themselves in both books. The process *is* analysis of variance, and so there need be no doubt about the name employed.

Statistical Methods: Applied to Experiments in Agriculture and Biology, Fourth Edition. *George W. Snedecor.* (Director of the Statistical Laboratory, Iowa State College; Head of the Statistical Section, Iowa Agricultural Experiment Station). Ames, Iowa: The Collegiate Press, Inc., 1946. Pp. xvi, 485. $4.50. *Two reviews follow:*

REVIEW BY D. J. FINNEY
Lecturer in the Design and Analysis of Scientific Experiment
University of Oxford, England

"DES PREUVES! sans doute il est bon d'avoir des preuves, mais il est peut-être meilleur de n'en avoir pas,"* said General Greatauk of a famous Penguinian trial. Herein lies the great dilemma for the author of a book on applied statistics: he must avoid mathematical proofs, yet to convince his readers he must justify his procedure logically rather than dogmatically. Professor Snedecor has overcome the difficulty very successfully by careful choice of numerical examples and by empirical construction of the probability distributions required in elementary statistical tests. The nonmathematical student should gain in understanding of the tests very considerably by repeating his sampling experiments. Since its first publication in 1937, this book has been one of the few to combine successfully a sound theoretical basis with an exposition sufficiently clear and detailed for those without statistical experience. The rewritten and greatly improved fourth edition contains 60 pages more than the third and has a slightly smaller but better type; it is a valuable text for the experimentalist in any branch of biology, whether he be a novice or experienced in statistics, and no criticism of detail should be considered as disparaging to the whole.

In chapter 1, the concepts of population and sample parameters, null hypotheses, tests of significance, and fiducial limits are introduced through sampling experiments on a table of random numbers. Applications of the binomial distribution are discussed and a useful table of fiducial limits for binomial sampling is given. Professor Snedecor's insistence on the importance

* Proofs! undoubtedly it is good to have proofs, but it is perhaps better not to have them.

of problems of estimation is a valuable corrective to the absurdity of regarding tests of significance as the sole function of statistical analysis. The presentation of the logic of the fiducial argument, however, seems open to criticism in some places, since it may suggest to the reader that a probability of the population parameter lying between certain limits is being determined. In fact, the probability is that of obtaining a sample equally or more extreme from a population whose parameter is specified. Perhaps the reviewer attaches undue importance to the distinction between these two forms of statement, but he is convinced that the student who learns to appreciate this point will be saved from many future illogical arguments.

Continuous distributions are introduced in chapter 2 and the ideas of chapter 1 are extended to them. Properties of normal distributions are investigated in chapter 3 with the aid of repeated sampling from a typical distribution, and the relationships of sample mean, sample variance, and the statistic t are discussed. In chapter 4, tests of difference between two samples are described. The approximate methods proposed for use when the two samples have unequal variances are likely to be sufficiently good for most occasions in which this difficulty arises, but Fisher and Yates' tables, which make the Fisher-Behrens test so easy to apply despite its logical subtleties, might well have been mentioned. The short chapter on computing methods contains many useful hints for the novice and rightly emphasizes the value of common sense appreciation of data in guarding against gross arithmetical errors.

Linear regression is introduced with a very clear explanation of the meaning of the different components of the total sum of squares, though the vague statement that the standard deviation from the regression line is "an average of the vertical distances from the line" might well be amended. Correlation is discussed in the next chapter and its relationship to regression is emphasized. The author fails to make clear that the tests of significance of departure from zero for the linear regression coefficient and the correlation coefficient are identical, and suggests a distinction between the validity of the two; the test of the correlation coefficient is not biased by selection or restriction of values for one variable, providing that the population regression is linear and the frequency distribution within arrays of the other variable is normal.

Chapter 8 describes "large sample" calculations for normal distributions. In chapter 9, the author returns to the analysis of enumeration data and gives a good account of the use of the χ^2 test for contingency tables. In particular, he describes the procedure for $2 \times 2 \times 2$ tables, a method not usually found in textbooks. Chapters 10 and 11 expound the analysis of variance. After a sampling experiment on the variance ratio distribution, the meaning of variance analysis, in terms of additive components of variance, is considered. Some elementary classes of experimental design are described, though the account of these could with advantage be extended so as to indicate the possibilities of incomplete block and lattice arrangements. Indeed, the reviewer would like to see chapter 11 made still longer in order to give a more comprehensive account of analysis of variance in relation to experimental

design. The statement that "one missing item can be supplied with the aid of a formula" should be given more precise phrasing as it tends to perpetuate the belief that loss of a few plots or animals matters little, since the statistician can always find out what the results would have been! The excellent account of modifications required for disproportionate subclass numbers has been much extended from the previous edition. Covariance analysis is well illustrated by typical examples in chapter 12.

In his account of multiple regression, Professor Snedecor pays more attention to correlation coefficients than is usual; the calculation of the inverse matrix directly from sums of squares and products is generally preferable to similar calculations from correlation coefficients. The subsequent account (chap. 14) of curvilinear regression begins with useful suggestions on how to decide the best form of function, and continues with illustrations of the fitting of polynomial regressions. Since published tables of orthogonal polynomial values are sufficient for most practical needs, their use might have been described rather than Fisher's original more complex procedure for fitting orthogonal polynomials to equally spaced observations; the latter method is now only required for observations of unequal weight, a complication beyond the scope of this book.

The account of individual degrees of freedom (chap. 15) is a useful introduction to the analysis of factorial experiments but scarcely does justice to the power of factorial design as an instrument of research; there is no mention of the possibility of reducing the errors of the more important treatment contrasts by confounding those of lesser interest. It is surprising to find a suggestion that Fisher and his associates have met with difficulty in maintaining the thesis that for each degree of freedom an appropriate sum of squares can be segregated! In chapter 16 further aspects of binomial and Poisson frequency distributions are considered, and transformations of data to a form more suitable for analysis of variance are discussed.

The final chapter summarizes the main principles of the planning and analysis of sample surveys and is an excellent introduction to a most important branch of applied statistics which is being increasingly used to day in many fields of investigation. This chapter will be difficult reading for those making their first acquaintance with statistical methods, and possibly in subsequent editions Professor Snedecor might consider its amplification.

The book contains all the usual tables now customary in textbooks of statistics for biologists, together with several less commonly found, of which the table of confidence intervals for the binomial distribution and the short table of range in normal samples are likely to be the most useful.

The reviewer has noticed the following misprints, most of which are trivial: p. 126, l. 6 from bottom: *for* 242% *read* 2421%; p. 221, l. 16 from bottom: *for* 173 *read* 172; p. 231, l. 7 above table: *for* 60.1 *read* 60.4, *for* 3.55 *read* 3.56, l. 5 above table: *for* 60.1 *read* 60.4; p. 245, l. 2 from bottom: *for* 3 points, (*a*, *b*), (*a*, *c*), and (*b*, *c*) *read* 6 points, (*a*, *b*), (*b*, *a*), (*a*, *c*), (*c*, *a*), (*b*, *c*), and (*c*, *b*); p. 268, l. 20: *for* Allen *read* Allan; p. 290, note to Table 11.22: *for*

67,434 *read* 67.434; p. 295, l. 8: *for* 9.8 *read* 9.9; p. 312, l. 8: *for* $(r-1)(6-1)$ *read* $(4-1)(6-1)$; p. 315, l. 14 from bottom: *for* 8.7 *read* 8.8; p. 316, l. 15: *for* Allen *read* Allan; p. 336, last line of Example 12.10: *for* 15.13 *read* 15.12; p. 337, l. 8: *for* 11.10 *read* 11.11; p. 363, l. 11: *for* 0.7412 *read* 0.7411, l. 10 from bottom: *for* 0.7412 *read* 0.7411; p. 408, l. 2: *for* 12 *read* 12.13, l. 4: *for* 12 *read* 12.10; p. 413: Last two paragraphs refer to an analysis of variance in Table 15.2, but no analysis of variance was given there; p. 427, l. 2 above table: for q read k (*twice*); p. 477, Index: *for* Allen *read* Allan; p. 489, Index under "Variance homogeneity of": *for* 349 *read* 249.

REVIEW BY W. J. YOUDEN
Physical Chemist, Boyce Thompson Institute for Plant Research, Inc. Yonkers, New York

PROFESSOR SNEDECOR's text on statistical methods is written for people without mathematical background with the purpose of teaching them how to apply statistical methods to the data of biology and agriculture. Mathematical statisticians, who are responsible in the first place for establishing the distributions of various statistics, sometimes feel that it is a dangerous business to turn people loose, equipped with a collection of formulas and with only a very imperfect understanding of the theoretical foundations of the subject matter of statistics. Certainly, scientific literature turns up an annual crop of inept applications of statistical techniques with a high yield of erroneous and even absurd conclusions. The reviewer is convinced that many of these mistakes might be avoided if the experimenter did not so often part company with intimate knowledge of his material when he applies a statistical formula. It seems obvious enough that when the results of a statistical analysis are in conflict with an interpretation based upon long experience and familiarity with similar data it is time to show caution and examine carefully the appropriateness of the statistical method used.

This particular hazard is not peculiar to statistics. Similar dangers exist in the use of chemical techniques. It is impossible for nonchemists to avoid chemical procedures; and since they are not chemists, they sometimes make chemical blunders. It may be hoped that Professor Snedecor's hint to seek expert advice when in doubt will be followed whether it be in the field of statistics, chemistry, or something else.

The first chapter, "Sampling of Attributes," introduces some fundamental notions by extended exercises in sampling from a table of random numbers. These samples of a binomial distribution are used to establish experimentally the frequency distribution for chi-square, and this is in turn compared with the exact distribution as a means of establishing confidence in the distributions derived mathematically. The discussion of the term "unbiased" might have made clear that it is used not only with respect to a sample but also to a statistic which is used to estimate a parameter of the population. There is a misprint in the formula at the foot of page 26 where $(1-r)$ is printed for $(1+r)$. One other misprint was noted on page 183, example 8.27, which gives

the equation for the normal curve. The exponent is misplaced so that it appears to be a part of the upper limit attached to the integral sign.

Great care is taken in the early chapters to keep before the reader just what he is attempting to do. It is continually emphasized that an experimenter has some object in mind when he collects data and that the purpose of the statistician is to set up a clear-cut corresponding hypothesis with an appropriate statistical test. Throughout the book, the illustrative examples as well as the exercises are chosen from research data with citations to original sources. This feature greatly enhances the use of the book by research workers who must dig out the methods without benefit of classroom instruction.

After discussing comparisons between two groups, linear regression, and correlation, there is a lengthy and excellent presentation of the analysis of variance and covariance. The remainder of the text covers multiple and curvilinear regression, individual degrees of freedom, a brief treatment of the binomial and Poisson distributions and concludes with an interesting chapter on the design and analysis of sampling which has been added since the first edition. In all, about 150 pages have been added since the first edition. Certain chapters and parts of chapters are indicated as forming a short introductory course in the elements of statistics. The necessary mathematical tables are scattered throughout the book but may be readily found by consulting an index of tables at the back of the book. It is especially encouraging to find how much attention is given to the vicissitudes of actual data. Such difficulties as unequal numbers of items in the groups compared, missing values from otherwise well-planned experimental data, unequal variances in the groups of data under comparison, and transformations of data to bring them into a normal distribution arise continually in actual data and are often ignored in elementary texts.

In conclusion, the question as to whether it is wise for experimental scientists to make elementary statistical applications without possessing a background in mathematical statistics must be passed upon by the scientists themselves. It is evident that ever larger numbers of scientists do find some familiarity with statistics of increasing value. Professor Snedecor's book has already earned the gratitude of many research workers and the new edition, with its inclusion of recent developments in statistics, should get a cordial reception.

Lectures on Statistical Methods of Inspecting and Controlling Quality: Delivered During a Course of Instruction Arranged by the Department of Munitions for Members of the Technical Staff: Held at the University of Melbourne, August 1944. Commonwealth of Australia, Ministry of Munitions, Quality Control Lectures, Melbourne, 1944. Edited by *W. N. W. Wallace* (Head of Quality Control Department, Directorate of Explosives Supply). Lectures by *M. H. Belz* (Associate Professor of Mathematics, University of Melbourne), *J. Churcher* (Engineer, Ammunition Factory, Footscray), *E. A. Cornish* (Officer in Charge, Section of Mathematical Statistics, Council for Scientific and Industrial Research),

H. S. Dean (Assistant Manager, Ammunition Factory, Finsbury), *C. S. Kneebone* (Quality Control Officer, Explosives Factory, Salisbury), *R.G. Moore*(Works Manager, Ammunition Factory, Footscray), *A. L. Stewart* (Standards Association of Australia), *R. S. Thompson* (Engineer, Ammunition Factory, Footscray), *Major Tillensor* (Army Inspection Service) and *W. N. W. Wallace*. East Melbourne S.2, Australia; Council for Scientific and Industrial Research (314 Albert St.), 1944. Pp. 236. Paper, mimeographed.

REVIEW BY EDWIN G. OLDS
Associate Professor of Mathematics
Carnegie Institute of Technology

EARLY in 1943, Mr. A. L. Stewart of the Standards Association of Australia, and Mr. R. G. Moore, Works Manager of the ammunition factory at Footscray, were sent to America by the Australian Department of Munitions to study the application of Statistical Quality Control. While in this country, they conferred with leaders in the field and visited a number of companies where quality control was in operation. Mr. Stewart attended several of the intensive eight-day courses sponsored by the Office of Production Research and Development of the War Production Board, and Mr. Moore visited at least one such course.

Impressed by our success in using statistical methods to improve the quality and increase the production of war material, they returned home convinced that education in the techniques should be made available to Australian manufacturers. Accordingly, an intensive course of instruction was scheduled at the University of Melbourne, August 9–16, 1944, by the Department of Munitions, for those members of its technical staff responsible for the quality of production. A total of eight-seven men attended the sessions held on the opening day and forty-seven of these remained for the entire course.

The book under review is a mimeographed report of the intensive course of lectures. It is assumed that a stenographic record of the proceedings was made and the result carefully edited. The copy read by this reviewer was struck off from badly-worn stencils, and is illegible in spots, but in so far as it can be read, the book has a surprisingly small number of minor errors. Most of these have been noted on a correction sheet at the end of the book.

The reviewer found the book tremendously interesting, and believes his interest will be shared by a great many of the nearly two thousand men who attended O.P.R.D. Quality Control courses in the United States during the war period. Moore and Stewart left the United States while our course was in an early stage, yet the further developments in the two countries followed surprisingly parallel lines. Students of quality control who have been wishing for a complete record of the O.P.R.D. course will find this book an excellent substitute and, in many instances, will have considerable difficulty in convincing themselves that the lectures they are reading are not the ones to which they themselves listened. Other persons will find the book a valuable introduction to the field.

The course at Melbourne was conducted by five instructors, (including Moore and Stewart) and seven special lecturers. After an opening day, arranged to give executives an overview of statistical quality control and its value as a production tool, the class was taught how to construct control charts for measured characteristics, fraction defective, and defects per unit. Lectures on the normal curve, binomial distribution and Poisson distribution, supplied the theory necessary for control charts, and paved the way for later work on acceptance sampling. Reports by the special lecturers served to give the students views of quality control at work.

A considerable amount of attention was given to the use of Sealy's modified control limits, and his method for controlling a group of machines was presented. The distinction between standard and modified limits was carefully made and factors dictating the choice of the latter were noted.

Following a discussion of the general problem of acceptance sampling by attributes, the students were given a thorough explanation of the Dodge-Romig Tables. In the main, this part of the instruction followed closely the material in the three papers originally published in the *Bell System Technical Journal*,* and later issued in book form,† by John Wiley and Sons. Acceptance sampling by variables also received attention.

In comparison with the O.P.R.D. courses, the Melbourne course seems to have devoted considerably less time to demonstration lectures, and to laboratory work by the students. On the other hand, somewhat more theory was included, and such subjects as Fisher's t-test, the estimate of sigma from successive differences, testing the reliability of chemical analyses, and rational sentencing of ammunition by use of 5-defect samples were included. Otherwise, the courses were very similar, relying on much the same reference material and using about the same order of presentation of topics. Even the reports on applications by special lecturers sound very familiar. For example, trouble with 100 per cent inspection at one plant was traced to the viewers (inspectors) who "were doing the job to the best of their ability" but whose average age was "of the order of 75 years."

The copy of the book used for this review bears the stamp of the Australian Legation, Scientific Research Liaison Office, 1910 K Street, N.W., Washington 6, D.C. Information is not available here regarding the supply available for distribution, or the possibility that a printed edition may be in preparation. If obtainable, the book will be fascinating to "alumni" of the O.P.R.D. courses, and should be stimulating reading for anyone who wishes to deepen his understanding of the methods and applications of statistical quality control.

* H. F. Dodge and H. G. Romig, "A Method of Sampling Inspection," *Bell System Technical Journal* 8: 613–31 October 1929. H. F. Dodge and H. G. Romig, "Single Sampling and Double Sampling Inspection Tables," *Bell System Technical Journal* 20: 1–61 January 1941. D. B. Keeling and L. E. Cisne, "Using Double Sampling in a Manufacturing Plant," *Bell System Technical Journal* 21: 37–50, June 1942.

† H. F. Dodge and H. G. Romig, *Sampling Inspection Tables*, New York: John Wiley & Sons, Inc., 1944. Pp. vi, 106. *Reviewed September 1945.*

PUBLICATIONS RECEIVED

Statistical methodology publications received are listed elsewhere in this issue under the section "Statistical Methodology Index."

Interviewing for NORC. Denver 10, Colo.: National Opinion Research Center, University of Denver, 1945. Pp. ix, 154. $2.00.

Revenues and Expenses of Newspaper Publishers in 1941: Together With Data Relating to National Advertising in Newspapers. *Neil H. Borden, Malcolm D. Taylor,* and *Richard T. Hovde.* Boston, Mass.: Division of Research, Harvard Business School, 1946. Pp. 56. $0.50. Paper.

Georgia Facts in Figures: A Source Book. By the *Citizens Fact-Finding Movement of Georgia.* Athens, Ga.: University of Georgia Press, 1946. $2.00. Paper.

Jobs and Markets: How to Prevent Inflation and Depression in the Transition. *Melvin G. DeChazeau, Albert G. Hart, Gardiner C. Means, Howard B. Myers, Herbert Stein,* and *Theodore O. Yntema.* Committee for Economic Development Research Study. New York: McGraw-Hill Book Co., Inc., 1946. Pp. xi, 143. $1.60.

Proceedings of the Institute of Economics and Finance, Occidental College, Los Angeles: Fourth Session, February 28, March 1, 2, 1945. Edited by *Cecil L. Dunn* and *Raymond G. McKelvey.* Los Angeles 41, Calif.: Institute of Economics and Finance, Occidental College, 1945. Pp. 161. Paper.

Income From Independent Professional Practice. *Milton Friedman* and *Simon Kuznets.* Publications of the National Bureau of Economic Research, Inc., No. 45. New York 23: National Bureau of Economic Research, Inc. (1819 Broadway), 1945. Pp. xxxiii, 594. $4.50. *To be reviewed.*

Should State Unemployment Insurance Be Federalized? *Herman Gray.* New York 17: *American Enterprise Association, Inc.* (4 East 41st St.), 1946. Pp. 71. $0.50. Paper.

Detention and Prosecution of Children: Jail Detention and Criminal Prosecution of Children of Juvenile Court Age in Cook County, 1938–1942. *Fred Gross.* Chicago 5, Ill.: Central Howard Association, 1946. Pp. xix, 177. $1.50. Paper.

Postwar Taxation and Economic Progress. *Harold M. Groves.* Committee for Economic Development Research Study. New York: McGraw-Hill Book Co., Inc., 1946. Pp. xiv, 432. $4.50.

Mathematics of Investment, Third Edition. *William L. Hart.* Bound with *Tables for Mathematics of Investment,* Third Edition. Boston, Mass.: D. C. Heath & Co., 1946. Pp. vii, 304; ii, 126. $3.60. *To be reviewed.*

Depreciation Policy and Postwar Expansion. *Lewis H. Kimmel.* Washington: Brookings Institution, 1946. Pp. vi, 66. $0.50. Paper.

The Economics of Housing: As Presented by Economists, Appraisers, and Other Evaluating Groups. *Laura M. Kingsbury.* New York 27: King's Crown Press (2960 Broadway), 1946. Pp. xii, 177. $2.50. Paper.

Commercial Policy in the Canadian Economy. *Orville John McDiarmid.* Cambridge: Harvard University Press, 1946. Pp. xii, 397. $4.50.

The Fundamentals of Accounting: A Cost and Revenue Approach. *Donald H. MacKenzie.* New York: Macmillan Co., 1946. Pp. xiii, 683. $4.00.

The Industrialization of Backward Areas. *K. Mandelbaum.* Oxford University, Institute of Statistics, Monograph No. 2. Oxford, England: Basil Blackwell, 1945. Pp. viii, 111. 10s. 6d.

Studies in the Classical Theories of Money. *Karl H. Niebyl.* New York: Columbia University Press, 1946. Pp. xiv, 190. $2.50.

In Search of the Regional Balance of America. Edited by *Howard W. Odum* and *Katharine Jocher.* Chapel Hill, N. C.: University of North Carolina Press, 1945. Pp. vii, 162. $3.00.

Small and Big Business: Economic Problems of the Size of Firms. *Joseph Steindl.* Oxford University, Institute of Statistics, Monograph No. 1. Oxford, England: Basil Blackwell, 1945. Pp. v, 66. Paper, 5s.; cloth, 7s. 6d.

The Theory of Price. *George J. Stigler.* An enlarged edition of *The Theory of Competitive Price,* 1942. New York: Macmillan Co., 1946. Pp. vii, 340. $3.75.

STATISTICAL METHODOLOGY INDEX, NO. 4

A QUARTERLY GUIDE TO CURRENT LITERATURE

Edited by

OSCAR KRISEN BUROS
Rutgers University

This bibliographic service presents statistical methodology literature (articles, books, theses, and chapters) published in 1945 to date. Anonymous references are listed first; other references follow in alphabetical sequence by first-named authors. Volume number, issue number (in parentheses), pagination, and date of issue are given for each journal reference. Stars indicate books, theses, and pamphlets; asterisks indicate publications seen by the editor of this INDEX. *Authors of papers on statistical methodology are invited to send reprints to Oscar K. Buros, Rutgers University, New Brunswick, New Jersey in order to facilitate prompt and accurate listings. Foreign-language papers should be accompanied by English translations of titles. Information is desired about references omitted.*

Professor P. C. Mahalanobis. *Sci & Cul* 10(10):422-3 Ap '45.* [473

Sample Surveys. *Economist* 150(5352): 445-6 Mr 23 '46.* [474

Society of Quality Control Engineers in Michigan. *Ind Qual Control* 2(5):15-6 Mr '46.* [475

Statistical Methods in Quality Control: 10, Classification of Defects and Quality Rating. American Institute of Electrical Engineers, Subcommittee on Educational Activities (J. Manuele, H. F. Dodge, A. I. Peterson, and R. E. Wareham). *Elec Eng* 65(3):117-9 Mr '46.* [476

Practical Statistics for Practical Physicists. Ralph Hoyt Bacon (Fairchild Camera and Instrument Corp., Jamaica N. Y.). *Am J Physics* 14(2):84-98 Mr-Ap '46.* [477

Analysis of Scores From Smelling Tests. William Dowell Baten (Professor of Mathematics, Michigan State College). *Biometrics B* 2(1):11-4 F '46.* [478

Some Comments on Carnap's Logic of Induction. Gustav Bergmann (Assistant Professor of Philosophy, State University of Iowa). *Philos Sci* 13(1):71-8 Ja '46.* See 317. [479

Approximation of Chi-Square by "Probits" and by "Logits." Joseph Berkson (Division of Biometry and Medical Statistics, Mayo Clinic, Rochester, Minn.). *J Am Stat Assn* 41(233):70-4 Mr '46.* [480

Rapid Least-Squares Solution of Polynomials. Abstract. R. T. Birge (Chairman, Department of Physics) and J. W. Weinberg (Instructor in Physics). (University of California, Berkeley.) *Physical R* 69(3-4): 106 Ag 1-15 '45.* [481

Expectancy Tables: A Method of Interpreting Correlation Coefficients. Reign H. Bittner (Owens-Illinois Glass Company, Toledo, Ohio) and Carlton E. Wilder (Captain, A.G.D., Northington General Hospital, A.S.F.). *J Exp Ed* 14(3):245-52 Mr '46.* [482

On an Equation of Wald. David Blackwell (Assistant Professor of Mathematics, Howard University). *Ann Math Stat* 17(1):84-7 Mr '46.* [483

Statistical Studies on New Zealand Wheat Trials: I, The Efficiency of Lattice Designs; II, The Analysis of Lattice Trials With Incomplete Data. S. W. Boyce (Wheat Research Institute, Department of Scientific and Industrial Research, Wellington, N. Z.). *N Z J Sci & Tech, Sect A* 27(3): 270-5, 276-80 O '45.* [484

Some Preliminary Problems of Sample Design for a Survey of Retail Trade Flow. Ralph F. Breyer (Associate Professor of Marketing, University of Pennsylvania). *J Marketing* 10(4):343-53 Ap '46.* [485

On the Interpretation of the Correlation Coefficient as a Measure of Predictive Efficiency. Hubert E. Brogden (Personnel Research Section, The Adjutant General's Office, New York). *J Ed Psychol* 37(2): 65-76 F '46.* [486

Some Distributions of Sample Means. George W. Brown (Research Engineer, R.C.A. Laboratories, Princeton, N. J.) and John W. Tukey (Assistant Professor of Mathematics, Princeton University). *Ann*

Math Stat 17(1):1–12 Mr '46.* Abstract: *B Am Math Soc* 52(3):242 Mr '46.* [487

Statistical Methology Index No. 3: A Quarterly Guide to the Current Literature. Edited by Oscar Krisen Buros (Director, Institute of Mental Measurements, Rutgers University). *J Am Stat Assn* 41(233): 144–54 Mr '46.* [488

Statistical Sampling Techniques and Marketing Research. Ralph Cassady, Jr. (Professor of Marketing, University of California, Los Angeles). *J Marketing* 9(4):317–41 Ap '45.* [489

Chains of Rare Events. Felix Cernuschi and Louis Castagnetto. (Harvard University.) *Ann Math Stat* 17(1):53–61 Mr '46.* [490

Philosophical Aspect of Statistical Theory. C. West Churchman (Lecturer on Philosophy, University of Pennsylvania). *Philos R* 55(1):81–7 Ja '46.* [491

Graduate Training in Statistics. W. G. Cochran (Institute of Statistics, University of North Carolina, Raleigh). *Am Math Mo* 53(4):193–9 Ap '46.* [492

Statistics in Pictures Sell Safety to Supervisors. J. F. Collins (Supervisor of Safety, Youngstown Sheet & Tube Co., Youngstown, Ohio). *Factory Mgmt* 104(4): 75–7 Ap '46.* [493

★**Methods in Climatology.** Victor Conrad (Research Associate in Climatology, Harvard University). Cambridge, Mass.: Harvard University Press, 1944. Pp. xx, 228. $4.00. (London: Oxford University Press, 1945. 22s. 6d.) *Reviewed in this issue.* [494

How to Use the Control Chart in School Management. Francis G. Cornell (Chief, Statistical Research Service, U. S. Office of Education, Washington) and Charles A. Bicking (Hercules Powder Co., Wilmington, Del.). *Sch Mgmt* 15(7):362+ Mr '46.* [495

The Estimation of Variance Components in Analysis of Variance. S. Lee Crump (Statistical Laboratory, Iowa State College). *Biometrics B* 2(1):7–11 F '46.* [496

A Note on Some Single Sampling Plans Requiring the Inspection of a Small Number of Items. J. H. Curtiss (Assistant to the Director, Bureau of Standards, Washington). *Ann Math Stat* 17(1):62–70 Mr '46.* [497

On the Use of the Sample Range in an Analogue of Student's *t*-Test. Joseph F. Daly (Lt. Comdr., USNR, Bureau of Ships, Washington). *Ann Math Stat* 17(1):71–4 Mr '46.* [498

On the Design of a Sample for Dealers' Inventories. W. Edwards Deming (Adviser in Sampling, Bureau of the Budget, Washington) and Willard Simmons (U. S. Office of Education, Washington). *J Am Stat Assn* 41(233):16–33 Mr '46.* [499

Suggestions With Respect to Experimentation Under School Conditions. Max D. Engelhart (Board of Examinations, Chicago City Junior College). *J Exp Ed* 14(3): 225–44 Mr '46.* [500

The U. S. Army Quartermaster Corps' Use of Sequential Sampling Inspection. Norbert L. Enrick (Formerly with Headquarters, QMC Inspection Service). *Ind Qual Control* 2(5):12–3 Mr '46.* [501

Historical Note on the Purchasing Power Concept, and Index Numbers. Wirth F. Ferger (Assistant Head, Business and Industrial Research Division, Bureau of Internal Revenue, Washington). *J Am Stat Assn* 41(233):53–7 Mr '46.* [502

Field Sampling for the Estimation of Wireworm Populations. D. J. Finney (Lecturer in the Design and Analysis of Scientific Experiment, University of Oxford). *Biometrics B* 2(1):1–7 F '46.* [503

Recent Developments in the Design of Field Experiments: I, Split-Plot Confounding; II, Unbalanced Split-Plot Confounding. D. J. Finney. *J Agric Sci* 36(1):56–62, 63–8 Ja '46.* [504

The Safety of Structures. Alfred M. Freudenthal (Lecturer in Bridge Engineering, Hebrew Institute of Technology, Haifa, Palestine). *Proc Am Soc Civil Eng* 71(8): 1157–91 O '45.* Discussion: 72(1, 2):111–4, 251–5 Ja, F '46.* [505

Measures of Heterogeneity in Agricultural and Similar Fields, and Their Inter-Relations. Birendranath Ghosh (Statistical Laboratory, Presidency College, Calcutta, India). *Sci & Cul* 11(7):382–3 Ja '46.* [506

An Estimate of the Correlation Coefficient of a Bivariate Normal Population When *X* Is Truncated and *Y* Is Dichotomized. Leonard Gillman and Harry H. Goode. *Harvard Ed R* 16(1):52–5 w '46.* [507

Unbiased Estimates for Certain Binomial Sampling Problems With Applications. M. A. Girshick (Principal Statistician, U. S. Department of Agriculture), Frederick Mosteller (Research Mathematician, Statistical Research Group, Princeton University), and L. J. Savage (Research Mathematician, Statistical Research Group, Columbia University). *Ann Math Stat* 17(1):13–23 Mr '46.* [508

Random Distributing of Lines in a Plane. S. Goudsmit (Professor of Physics, University of Michigan). *R Mod Physics* 17(2–3): 321–2 Ap-Jl '45.* [509

On the Independence of the Extremes in a Sample. E. J. Gumbel (Visiting Professor,

New School for Social Research). *Ann Math Stat* 17(1):78–81 Mr '46.* [510

The Theory of Unbiased Estimation. Paul R. Halmos (Assistant Professor of Mathematics, Syracuse University). *Ann Math Stat* 17(1):34–43 Mr '46.* [511

★Notes on Precision of Sample Estimates: Technical Notes on the Formulas Used to Evaluate the Precision of Data. Morris H. Hansen (Statistical Assistant to the Director), William N. Hurwitz, (Chief, Sampling Section, Office of the Director), and Irving Holmes (Assistant Chief, Agriculture Division). (Bureau of the Census, Washington.) Washington: U. S. Government Printing Office, 1945. Pp. 283–91. $0.10. Paper.* [512

What Does Management Expect From Quality Control? Myron J. Hayes (Eastman Kodak Co., Rochester, N. Y.). *Ind Qual Control* 2(5):8–9+ Mr '46.* [513

Symmetrical Incomplete Block Designs With $\lambda=2, k=8$ or 9. Q. M. Hussain (Statistical Laboratory, Presidency College, Calcutta, India). *B Calcutta Math Soc* 37(3):115–23 S '45.* [514

On the Characteristic Function of the Distribution of the Product of Two Normal Variates. J. O. Irwin (London School of Hygiene and Tropical Medicine, London). *Proc Cambridge Philos Soc* 42(1):82–4 F '46.* [515

Serial Correlation. Nathan Jaspen (United States Employment Service). *Psychometrika* 11(1):23–30 Mr '46.* [516

Certain Properties of the Correlation Coefficient. H. G. Johnson (University of Minnesota). *J Exp Ed* 14(3):263–6 Mr '46.* [517

On the Distribution of Values of Sums of the Type $\Sigma f(2^k t)$. M. Kac (Assistant Professor of Mathematics, Cornell University). *Ann Math* 47(1):33–49 Ja '46.* [518

On the Selection of an Inspection Plan. A. J. Kavanagh (Scientific Instrument Division, American Optical Co.). *Ind Qual Control* 2(5):10–1 Mr '46.* [519

★Statistische Methoden für Naturwissenschafter, Mediziner, und Ingenieure. [Statistical Methods for Natural Scientists, Medical Research Workers, and Engineers.] Arthur Linder (Professor of Applied Mathematical Statistics, University of Geneva; Instructor at the Cantonal Technical University in Zurich). Basel, Switzerland: E. Birkhäuser & Co., 1945. Pp. 150.* Cloth, Frs. 18.50; paper, Frs. 15.50. *Reviewed in this issue.* [520

Correction to the Paper "On a Problem of Estimation Occurring in Public Opinion Polls." H. B. Mann (Associate Professor of Mathematics, Ohio State University). *Ann Math Stat* 17(1):87–8 Mr '46.* [521

Neumann's and Morgenstern's New Approach to Static Economics. J. Marschak (Professor of Economics, University of Chicago). *J Pol Econ* 54(2):97–115 Ap '46.* [522

Distribution of Sample Arrangements for Runs Up and Down. P. S. Olmstead (Consultant in Engineering Statistics, Bell Telephone Laboratories, Inc., Murray Hill, N. J.). *Ann Math Stat* 17(1):24–33 Mr '46.* [523

A Note on Sampling Inspection. Paul Peach (Industrial Statistician, Institute of Statistics, University of North Carolina, Raleigh) and S. B. Littauer (Professor of Mathematics, Newark College of Engineering). *Ann Math Stat* 17(1):81–4 Mr '46.* [524

A New Descriptive Statistic: The Parabolic Correlation Coefficient. Charles C. Peters (Emeritus Professor of Education, Pennsylvania State College). *Psychometrika* 11(1):57–69 Mr '46.* [525

Can an Advertiser Believe What Surveys Tell Him? Alfred Politz. *Ptr Ink* 215(1):23–5 Ap 5 '46.* [526

Distributions Derived From the Multinomial Expansion. G. B. Price (Professor of Mathematics, University of Kansas). *Am Math Mo* 53(2):59–74 F '46.* [527

Information and the Accuracy Attainable in the Estimation of Statistical Parameters. C. Radhakrishna Rao (Statistical Laboratory, Presidency College, Calcutta, India). *B Calcutta Math Soc* 37(3):81–91 S '45.* [528

Use of Statistical Methods in Quality Control Program. W. T. Rogers (Metallurgical Statistician, National Tube Co., Lorain, Ohio). *Steel* 118(6):119–20 F 11 '46.* [529

Preparation of Graphs for Physical Papers. Duane Roller (Editor, American Journal of Physics; Professor of Physics, Wabash College). *Am J Physics* 14(2):133–5 Mr-Ap '46.* [530

★A First Guide to Quality Control for Engineers. E. H. Sealy (Ministry of Supply Advisory Service on Statistical Method and Quality Control, London). A reprint of the 1943 edition published by the Ministry of Supply. London: H. M. Stationery Office, 1945. Pp. 38. 1s. Paper. (New York 20: Sales Department, British Information Services, 30 Rockefeller Plaza. $0.30.)* *Reviewed September 1946.* [531

The Advancing Statistical Front. Walter A. Shewhart (Research Engineer, Bell Telephone Laboratories, Inc., Murray Hill,

N. J.). *J Am Stat Assn* 41(233):1–15 Mr '46.* [532

We Recommend the Teaching of Statistics in High School. Ollie Sanders Smith (Hartsville High School, Hartsville, S. C.) and others. *Math Teach* 39(4):182–3 Ap '46.* [533

Binomial Coefficients: Derivation of a Formula. Harry Sohon (Assistant Professor of Electrical Engineering, University of Pennsylvania). *J Eng Ed* 36(7):414–9 Mr '46.* [534

★**Sequential Analysis of Statistical Data: Applications.** Prepared by the Statistical Research Group, Columbia University for the Applied Mathematics Panel, National Defense Research Committee, Office of Scientific Research and Development. SRG Report 255, Revised; AMP Report 30.2R, Revised. New York: Columbia University Press, September 1945. Pp. vii, 17; iv, 80; v, 57; iii, 25; iii, 18; iii, 39; ii, 41. $6.25. (London: Oxford University Press, 1946. 42s.) *Reviewed March 1946.* [535

Graphical and Tabular Aids for Determining Sample Size When Planning Experiments Which Involve Comparisons of Percentages. Frances Swineford (Research Associate, Department of Education, University of Chicago). *Psychometrika* 11(1): 43–9 Mr '46.* [536

Multiple Regression for Systems of Equations. Gerhard Tintner (Associate Professor of Mathematics and Economics, Iowa State College). *Econometrica* 14(1): 5–36 Ja '46.* [537

A Note on Solutions of Analysis of Variance for the Problem of Unequal or Disproportionate Subclass Numbers. Fei Tsao and Palmer O. Johnson (Professor of Education, University of Minnesota). *J Exp Ed* 14(3):253–62 Mr '46.* [538

★**General Solution of the Analysis of Variance and Covariance in the Case of Unequal or Disproportionate Numbers of Observations in the Subclass Numbers.** Fei Tsao (National Central University, Chungking, China). Doctor's thesis, University of Minnesota, 1945. Pp. 115. [539

Maximum Validity of a Test With Equivalent Items. Ledyard R. Tucker (Head, Analysis and Research Unit, College Entrance Examination Board, Princeton, N. J.). *Psychometrika* 11(1):1–13 Mr '46.* [540

An Inequality for Deviations From Medians. John W. Tukey (Assistant Professor of Mathematics, Princeton University). *Ann Math Stat* 17(1):75–8 Mr '46.* [541

Incomplete Block Experimental Designs in Insect Population Problems. F. M. Wadley (Agriculture Research Administration, Bureau of Entomology and Plant Quarantine, Washington). *J Econ Entomol* 38(6):651–4 D '45.* [542

Some Significance Tests Based on Order Statistics. John E. Walsh (Department of Mathematics, Princeton University). *Ann Math Stat* 17(1):44–52 Mr '46.* [543

Probable Errors for Least-Squares Solutions of Polynomials. Abstract. J. W. Weinberg (Instructor in Physics) and R. T. Birge (Chairman, Department of Physics). (University of California, Berkeley.) *Physical R* 68(3–4):106 Ag 1–15 '45.* [544

Analytical Statistics in Industry. Holbrook Working (Professor of Prices and Statistics, Stanford University). *J Eng Ed* 36(7):407–13 Mr '46 * [545

On the Theory of the Brownian Motion II. Ming Chen Wang and G. E. Uhlenbeck. (H. M. Randall Laboratory of Physics, University of Michigan.) *R Mod Physics* 17(2–3):323–42 Ap-Jl '45.* [546

A Simple Graphical Method for Orthogonal Rotation of Axes. Wayne S. Zimmerman (University of Southern California). *Psychometrika* 11(1):51–5 Mr '46.* [547

Periodicals Represented for the First Time

B Calcutta Math Soc—*Bulletin of the Calcutta Mathematical Society.* 4 issues; Rs. 8 (Rs. 2); Calcutta Mathematical Society, 92 Upper Circular Road, Calcutta, India.

Harvard Ed R—*The Harvard Educational Review.* 4 issues; $2.50(75¢); Lawrence Hall, Kirkland St., Cambridge, Mass.

Iron & Steel Engr—*Iron and Steel Engineer.* 12 issues; $7.50($1.50); Empire Bldg., Pittsburgh, Pa.

J Agric Sci—*The Journal of Agricultural Science.* 4 issues; 30s.(10s.); Cambridge University Press, Bentley House, 200 Euston Road, London, N.W.1, England.

J Econ Entomol—*Journal of Economic Entomology.* 6 issues; $3(75¢); 450 Ahnaip St., Menasha, Wis.

J Eng Ed—*The Journal of Engineering Education.* 10 issues; $3(50¢); Prince and Lemon Sts., Lancaster, Pa.

J Pol Econ—*The Journal of Political Economy.* 6 issues; $5; University of Chicago, Chicago, Ill.

Math Teach—*The Mathematics Teacher.* 8 issues; $2(25¢); 525 West 120th St., New York 27, N. Y.

N Z J Sci & Tech, Sect A—*The New Zealand Journal of Science and Technology: A, Agricultural Section.* 6 issues; 10s.(2s.); The Government Printer, Wellington, N. Z.

Philos R—*The Philosophical Review.* 6 issues; $5($1); Cornell University Press, Ithaca, N. Y.

Physical R—*The Physical Review.* 12 issues; $15($1.50); American Institute of Physics, 57 East 55th St., New York 22, N. Y.

Proc Am Soc Civil Eng—*Proceedings of the American Society of Civil Engineers.* 10 issues; $8($1); 33 West 39th St., New York 18, N. Y.

Ptr Ink—*Printers' Ink.* 52 issues; $4; 205 East 42nd St., New York 17, N. Y.

R Mod Physics—*Reviews of Modern Physics.* 4 issues; $4($1.20); American Institute of Physics, 57 East 55th St., New York 22, N. Y.

Sch Mgmt—*School Management.* 11 issues; $1(20¢); 52 Vanderbilt Ave., New York 17, N. Y.

Sci & Cul—*Science and Culture.* 12 issues; Rs. 7/8 (Rs. 1/-); foreign, 15s.; Indian Science News Association, 92 Upper Circular Road, Calcutta, India.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 235 SEPTEMBER 1946 *Volume 41*

ONE STATISTICAL WORLD

EARL LATHAM
*University of Minnesota**

THE construction of a world political system has held the center of popular attention at the expense of other international developments of importance but the work of building international organizations also goes forward in many different subject matter and technical fields, less conspicuously and therefore less well known. While the dramatic issues of international politics command the stage, international institutions have been formed or are forming to deal with problems of health, agriculture, trade, finance, refugees, employment, education, and culture. These agencies of international administration share a common need for reliable data, scientifically collected and assembled, and disseminated as widely as demand requires.

Sound statistics are basic to the formulation and administration of international policies, and the organization of comprehensive statistical services goes apace with the development of the international agencies they will serve. A world system is emerging from numerous independent and incomplete fragments of statistical activity: the United Nations Statistical Commission; the Statistical Division of the Department of Economic Affairs of the UN; the specialized international agencies like the Food and Agriculture Organization; regional statistical organizations like the Inter American Statistical Institute; and private international statistical organizations like the International Statistical Institute. In addition to these world developments, parallel action is taking place within national statistical systems to order affairs in such a way that the responsibilities of membership in a world statistical system can be discharged most effectively. In the United States this parallel action is centered on a new organization called the Federal Committee on International Statistics. These efforts to create

* Formerly with the Division of Statistical Standards, Bureau of the Budget.

a world of statistics bear promise of creating new forms of professional cooperation which will reflect, institutionally, the true universality of statistical science.

The hub of this new statistical universe is the United Nations, with reference to which the various segmented and specialized parts should assume the ordered arrangement of wheel to axle. To use this figure of speech in another way, the skill with which axle and wheel are joined may determine both the rate and smoothness of passage of UN affairs, if not indeed their direction and course. Whether these affairs follow the track of Phaeton or Apollo will depend in great part upon the success with which international statistical service plots the course of policy that neither freezes nor burns.

UNITED NATIONS STATISTICAL COMMISSION

Of the five elements of international statistical organization mentioned, those of presiding importance are in the United Nations. The United Nations Statistical Commission comes first in point of time; and certainly not subordinate in measure of influence.

At the time of the San Francisco Conference, when the United Nations Charter was drafted, some effort was made to draw attention to the desirability of establishing international statistical services in the United Nations. A conference of the principal statistical officials in the United States Government was called by the Bureau of the Budget to prepare and agree upon a statement of need for international organization in March 1945, with representatives in attendance also from certain accessible international organizations, including the Economic, Financial, and Transit Department of the League of Nations, the International Labor Office, the UN Interim Commission on Food and Agriculture and the Inter American Statistical Institute. This conference proposed to the Secretary of State that questions relating to international statistical organization be considered at San Francisco but this was not done, and the United Nations Charter is not explicit on the subject.

The Preparatory Commission of the United Nations then met in London in the autumn of 1945 and the conference of international statisticians made further proposals to the Department of State which were transmitted with its approval to the United States Delegate to the Preparatory Commission. The United States Delegation did some excellent staff work in the preparation of a paper titled, "Observations on Organization of Statistical Work of the Secretariat."

It had previously been recommended that a Statistical Commission

be established in the United Nations Economic and Social Council. The United States delegation in its "Observations" proposed further that the statistical services of the United Nations be grouped within the UN Secretariat in an Office of Economic and Social Affairs, directly under an Assistant Secretary-General having charge of that office. As the Secretariat of the United Nations was finally developed, it was decided to create a Department of Economic Affairs and a Department of Social Affairs in the Secretariat instead of one combined department as had been recommended to the Preparatory Commission.

Heeding the recommendation of the Preparatory Commission, the Economic and Social Council in February 1946 established the Statistical Commission. By resolution adopted in June 1946 the Commission was charged with responsibilities for promoting the development of national statistics and improving their comparability; coordinating the statistical work of the specialized international agencies; developing the central statistical services of the Secretariat of the United Nations; promoting the improvement of statistics and statistical methods; and advising the organs of the United Nations on general questions relating to the collection, interpretation, and dissemination of statistical information. There is a close similarity between these functions and those vested in the Division of Statistical Standards of the Bureau of the Budget, a similarity which is not accidental since the example of the latter was freely used in the original draft statements of responsibility.

After establishing the Statistical Commission, the Economic and Social Council appointed a "nucleus commission" of nine members serving, not as nationals but as experts from the following countries: France, United Kingdom, India, United States, Norway, Brazil, China, the Ukrainian SSR and the USSR. The appointee from Brazil was not able to serve for reasons of ill health and the appointment reserved for the Ukraine was not acted upon. The nucleus commission met in New York in April and May of this year and elected as chairman its United States member, Dr. Stuart A. Rice, Assistant Director in charge of Statistical Standards, Bureau of the Budget. The Commission considered and reported to the Economic and Social Council on its definitive composition and terms of reference. All of its recommendations were adopted by the Council except that which proposed that Commission members serve in their private capacities; the Council resolved that members should be representatives of their respective governments.

In its permanent organization the Commission will contain not

more than twelve members, confirmed by the Economic and Social Council after nomination by member countries of the UN, and it will meet three times a year. It was the consensus of the Commission that regional statistical organizations should be encouraged and that the International Statistical Institute should be brought into some regular effective relation to the United Nations. Considerable attention was given to the statistical activities of the specialized international agencies and draft language was prepared for inclusion in the terms of agreement negotiated between the specialized agencies and the United Nations, covering the coordination of their statistical programs. The Commission voted to create a Sub-Commission on Sampling and the Council adopted its recommendation, authorizing the Commission itself to create the Sub-Commission, a departure from the ordinary procedure according to which the Council makes the appointments.

The Statistical Commission set the pace for other commissions of the United Nations in the speed and efficiency with which it set about the tasks of organizing itself, and marked out the lines for the development of the United Nations statistical services. It has the great advantage of being located near the top of the UN hierarchy and possessed of a separate organizational life and identity.

UNITED NATIONS SECRETARIAT

Since the Statistical Commission is an advisory body and since it will not be in continuous session, it is clear that the day to day tasks of international statistical administration will have to be performed by the "statistical secretariat" of the UN, a short-hand way of referring to the Statistical Division of the Department of Economic Affairs of the United Nations.

When the decision was made to create a Department of Economic Affairs and a Department of Social Affairs instead of a single Office of Economic and Social Affairs in the UN Secretariat, it then became necessary to decide in which of the two the central statistical office should be located. As between the two departments, the case for lodgement in Economic Affairs was the more persuasive since it is expected that the chief statistical problems will be economic. Although more persuasive, the case was nevertheless weak; the better arrangement would have been to establish a single Department of Economic and Social Affairs, since the distinction between the two subject matter fields is frequently an abstract and artificial one, and since the work of both Departments will be related to one organ, the Economic and Social Council.

The United States Delegation to the United Nations Preparatory

Commission in London proposed that the statistical secretariat be given substantial functions of control and coordination over statistical activities of the United Nations and its affiliates. It proposed that the secretariat be given responsibility for providing statistical information to the Secretary-General and the Assistant Secretary-General; validating statistical findings used in UN proceedings; performing functions of coordination of statistical activities within the UN and among the specialized international agencies; collecting statistics from member governments and other international organizations; maintaining certain statistical publications; collecting statistics, where feasible, for affiliated statistical organizations; recommending uniform standards for national and international statistics; consulting with and advising member governments and international organizations about statistical problems; and providing a general clearing house service for international statistical data. These recommendations were accepted by the Preparatory Commission as a valuable statement of the functions to be performed by the statistical secretariat.

At the moment of writing, the statistical secretariat has not been organized, but the general lines of its development and relationship to the Statistical Commission can be discerned. The Statistical Commission is responsible for recommending the general statistical policies of the United Nations; in doing so it will be assisted by the statistical secretariat. In addition, the continuous tasks of coordination of the statistical activities of the United Nations, its organs, and its affiliated specialized agencies will be performed by the secretariat. In a real sense therefore it will act as the central UN control in the development of systematic international statistical organization. For example, statistical data required by two or more international organizations may be collected by the central statistical secretariat, or may be collected by one of them designated by the secretariat on the recommendation of the Statistical Commission. It will probably have primary responsibility for collecting general purpose data of various kinds, like population and national income statistics, although it is doubtful whether it will actually gather much original data directly from respondents. It is more likely that its collecting function in the main will be one of assembling or collating data already in existence. There would, for example, be little point in having the statistical secretariat attempt to take the decennial census of the United States, even if it were constitutional or otherwise feasible for it to do so. Where general purpose statistical information is wanted from statistically developed countries, however, it is entirely possible that the central UN statistical

secretariat may undertake, with the cooperation of the country involved, to collect original data.

SPECIALIZED AGENCIES

The specialized international agencies are those that are or will be affiliated to the United Nations but not as an organic part of its structure. These organizations include: The International Bank for Economic Reconstruction and Development; the International Monetary Fund; the Food and Agriculture Organization; Provisional Civil Aviation Organization; United Nations Educational, Scientific, and Cultural Organization; and the International Labor Organization. In addition, the World Health Organization is being organized and plans are being made for an International Trade Organization and an International Refugee Organization. Each of these specialized agencies exists independent of the formal organization of the United Nations but will be related to it according to the terms of agreements of affiliation to be made with the United Nations. In their entirety they present special problems of statistical coordination for the Commission and for the Statistical Division of the Department of Economic Affairs.

The problem is not unlike that in the United States Government where numerous Federal agencies collect data in areas of their special competence and only the functions of coordination and the development of statistical standards are centralized. There is this difference, however: The central secretariat of the United Nations is expected to perform some functions of collection, as indicated. Otherwise the specialized agencies will pursue their own statistical programs, which will be harmonized with those of other such agencies and with the UN statistical secretariat, on the advice of the Statistical Commission. These relations will be defined in the terms of agreement of affiliation. The Statistical Commission is deeply concerned with the terms of these agreements of affiliation, since the opportunity is thus afforded to set out the basic lines along which the statistical activities of the specialized agencies may be coordinated.

The techniques of coordination are familiar, and presumably will be applied to this international statistical context. The basic pattern of relationships is decentralization. As already indicated, where two or more specialized agencies want the same kind of data, one of them may be designated as the focal agency with primary responsibility for its collection. Similarly other focal points may be designated for the receipt and distribution of data. Arrangements for the clearance of information about statistical activities can be operated through the

secretariat and the Statistical Commission. For example, the specialized agencies will be informed when there is any matter before the Commission in which they might be interested, so that they may attend and be heard. Their appearance at such deliberations is not a matter of right but of comity, since the Commission may want to discuss the statistical activities of the specialized agencies in executive session when some line of policy or decision is to be determined. More regular and routine arrangements for clearance can be managed through the secretariat, among such being the dissemination of reports and bulletins, circulation of announcements about events of interest and importance, meetings and conferences, and so on. A well devised reporting system that really communicates is an indispensable part of the clearance apparatus.

The uniform language covering the statistical activities of the specialized international agencies binds the UN and the agency to strive for maximum cooperation, the elimination of all undesirable duplication between them, and the most efficient use of their technical personnel in collection, analysis, publication, and dissemination of statistical information by each of them respectively. The agency and the UN also agree to combine their efforts to secure the greatest possible usefulness and utilization of statistical information and to minimize the burdens placed upon national governments and other organizations from which such information may be collected.[1]

The agency recognizes the United Nations as the central agency for the collection, analysis, publication, standardization and improvement of statistics serving the general purposes of international organizations. In turn, the United Nations recognizes the specialized international agency as the appropriate agency for the collection, analysis, publication, standardization, and improvement of statistics within its special sphere, without prejudice to the right of the United Nations to concern itself with such statistics so far as they may be essential for its own purposes or for the improvement of statistics throughout the world. In consultation with the specialized agencies the UN assumes the responsibility to develop administrative instruments and procedures through which effective statistical cooperation may be secured between the United Nations and the agencies brought into relationship with it. The terms of agreement also contain an affirmation that the collection of statistical information should not be duplicated by the

[1] See for example the "Draft Agreement Between the United Nations and the Food and Agriculture Organization of the United Nations," United Nations Economic and Social Council, Document E/57, June 10, 1946, Article XII, pp. 7–8.

United Nations or any of the specialized agencies whenever it is practicable for any of them to utilize information or materials which another may have available. In order to build up a central collection of statistical information for general use, it is agreed that data supplied to the agency for incorporation in its basic statistical series or special reports should so far as practicable be made available to the United Nations.

Some of the statistical needs of international administration have already been defined. For instance the Economic and Employment Commission has recommended to the Economic and Social Council that it will require firm data on national resources, human and material, such as plant capacity and labor force, if it is to do its work intelligently. The PICAO is holding conferences to determine how the cost of constructing thirteen weather stations in the North Atlantic should be distributed among the benefiting countries. In addition, it will need other statistical information on traffic, operations costs, and finance, if it is to do its work as an international agency effectively. It is certain that the World Health Organization will require numerous data on disease, nutrition, and medical services. The work of the Food and Agriculture Organization will require data on food supplies, crop acreage, food prices, crop forecasts, and agricultural methods. In more homely but very practical vein, the UN will need statistical information to fix the quotas of contribution that each nation will be expected to meet. If the UN salaries are paid in terms of net after taxes, systematic information on taxes throughout the world will have to be obtained, for UN recruits its personnel throughout the world. All of these statistical considerations and concerns grow out of the operations of international administration.

The Statistical Commission and the Statistical Division have an interest in all of these matters, as the list of delegated responsibilities discloses. The Commission will be interested in promoting the statistical development of such countries as China, for instance, in order to provide more reliable data on population. It may want to make recommendations looking towards improving the comparability of national income estimates for tax and quota purposes. The Statistical Division and the Commission would undoubtedly take steps to see that the World Health Organization and the FAO use the same data on nutrition, to take a random example of possible duplication because of related subject matter interests. The Commission will be interested in supporting the central secretariat in the development of special techniques for gathering and disseminating general purpose statistics. Of

use in this and other respects will be the Sub-Commission on Sampling which the Commission has been empowered to establish. At all points, the Commission will be in position to promote improved statistics and statistical methods by study and recommendation, and it will be available to the organs of the United Nations for consultation and advice on general questions of statistics.

THE LEAGUE OF NATIONS AND THE UN

These three elements—United Nations Secretariat, the Statistical Commission, and the specialized international agencies—are the three sides of the new international statistical center which is being formed. They represent a considerable advance over the statistical arrangements which prevailed under the League of Nations.[2] In the main the League's statistical work was based on the International Convention Relating to Economic Statistics drafted by the International Conference Relating to Economic Statistics which met in Geneva in 1928. Occupying the place now filled by the Statistical Commission was the Committee of Statistical Experts which was set up under the Convention to perform statistical services in major areas of economic interest. The Committee was assisted by the Secretariat of the League of Nations but the latter in no way dominated or even led the work of the Committee. It performed work on request of the League, its affiliates and on its own initiative.

The Committee of Statistical Experts, however, did not meet for the first time until March 1931, some twelve years after the establishment of the League of Nations. There was no meeting in 1932 but thereafter there were annual meetings until 1939, when the outbreak of war brought them to an end. The meetings were of short duration, lasting not more than eight days. The statistical work of the Committee was carried on by sub-committees at the time of the annual sessions and, in the interim, through correspondence. Under these conditions, studies were completed on a minimum list of countries, to be distinguished in foreign trade statistics; countries of provenance and destination in foreign trade; minimum list of commodities for international trade statistics; statistics of the gainfully occupied population; indices of industrial production; and on timber, housing, tourist and mining and metallurgical statistics. Some work was also done (the

[2] An account of the statistical and research activities of the League of Nations was prepared for the United Nations Statistical Commission by A. Rosenborg, Head of League of Nations Mission in the United States. This account draws attention to the study titled "The Economic and Financial Organization of the League of Nations, a Survey of Twenty-five Years Experience," written by Martin Hill, to be published shortly by the Carnegie Endowment for International Peace.

studies were never completed) on indexes of prices; financial statistics; balance of payments; and indices of quantum and prices in international trade. Other statistical activities of the League involved health and drug control, communications and transport, and territories and minority questions but were never very extensive in comparison with the work described.

Administrative coordination of all League statistical activity was operated through the Inter-Departmental Statistical Committee of the Secretariat, which also concerned itself with the standards of presentation of the relevant data published by the League, and which met at regular intervals under the chairmanship of the Head of the Economic Intelligence Service.[3] The secretary of the Committee of Statistical Experts served as the secretary of the Inter-Departmental Committee. The Chief of the Statistical Section of the International Labor Office participated in the meetings of the Committee which included within its purview questions of collaboration with the ILO. Other devices of coordination were the interlocking memberships of the ILO and the International Institute of Agriculture whose chiefs served *ex officio* as associated members of the Committee of Experts. Reciprocally, the Economic Intelligence Service was represented *ex officio* on the statistical expert committees of the ILO and the IIA.

Certain weaknesses in this scheme of arrangements have been repaired in the organization of the United Nations statistical services. First, the Statistical Commission will not do any of the actual work of collecting, assembling, publishing or disseminating statistics that was the responsibility of the Committee of Experts. This leaves the Commission free to devote itself to general problems of standards, policy and organization of statistical activity. Second, the Statistical Commission will meet three times as often as the Committee, assuring continuity of expert and disinterested attention to general policy matters. Third, the terms of relationship with the specialized international agencies, statistics-wise, fix in fundamental procedure the basic pattern of a comprehensive statistical service. Fourth, the organization of the United Nations' major statistical development begins with the beginning of the United Nations itself, and not twelve years later. Fifth, the central statistical organization of the United Nations is comprehensive in conception and plan; the statistical services of the League of Nations were not developed in the same degree. Sixth, to

[3] This paragraph is based on information contained in the account of A. Rosenborg, op. cit., and a related account of the League of Nations Committee of Statistical Experts, prepared by E. Dana Durand, U. S. Member of the Committee. Both these papers were submitted to the United Nations Statistical Commission for its consideration in April 1946.

the extent that the United Nations is politically stronger than the League (if it is), the statistical services of the United Nations are likely to be invigorated and fortified by the prestige and power of the parent authority.

REGIONAL STATISTICAL ORGANIZATIONS

One of the notable developments in statistical administration in the last six years is regional statistical organization, the furthest advanced and the most mature being the Inter American Statistical Institute, organized in 1940 as a result of conversations held at the Washington Session of the Eighth Scientific American Congress in that year. The Statistical Commission has recommended as a policy that regional statistical organizations be encouraged, largely as a result of the example which the Inter American Statistical Institute (called IASI) supplies. A brief account of the IASI may help to make clear what it is that the Statistical Commission sponsors by its approval.

The Twenty-fifth Session of the International Statistical Institute was originally scheduled for Washington in 1939. Upon the outbreak of war in Europe, the session was first postponed a year and then indefinitely deferred. At the suggestion of American members of the International Statistical Institute, a statistical section was added to the program of the Eighth American Scientific Congress. Members of the International Statistical Institute from four American nations met during the Congress to organize the IASI, with the principal objective of encouraging the development of statistical science and administration throughout the Western Hemisphere. In particular, its program was aimed at improving the methods used in the collection, tabulation, analysis, and publication of both official and unofficial statistics, and at obtaining a greater degree of international comparability in these statistics.

In the ensuing six years, the IASI has gained the official membership of almost all of the American governments, on whom it depends for support, and has come to play a very influential part in Western Hemisphere statistical affairs.[4] Consideration is being given to a pro-

[4] An illustration of this influence is the extent to which IASI members have been called upon to serve in various capacities in the *United Nations. Of the seventy-eight "constituent"* or professional members, three were designated as their governments' representatives in the first session of the eighteen-member Economic and Social Council of the UN. These included the Council's Second Vice President. Two others, the Institute's President and First Vice President, were appointed as members of the nine member "nucleus" Statistical Commission and one of these was elected by the Commission as its Chairman. Still another IASI member was recently designated by his government as the Ambassador of his country, Mexico, to the United States. Recommendations of the Institute's Executive Committee, meeting in Rio de Janeiro in January 1946, have already in numerous instances been given positive effect by various American governments.

posal to make the IASI the statistical arm of the Pan American Union, to make it, in short, an "operating" agency through which the collective statistical interests of the American governments can be implemented. If these negotiations fulfill the promise they bear, the IASI will become the official statistical instrument of the Inter American System, created by the Act of Chapultepec and tacitly recognized in Chapter VIII (on "Regional Arrangements") of the United Nations Charter.[5]

Interest in regional statistical organization is manifest in other parts of the world also. For instance, statistical leaders in India are exploring the possibility of creating a similar organization for India and the adjoining countries. During the years of the war, the Middle East Statistical Bureau with headquarters in Cairo did much to foster and develop improved statistical activity in the Middle East. In Europe, the Emergency Economic Commission for Europe sketched the outlines of European statistical organization. There is similar interest in regional statistical organizations in East Africa and South Africa. Except for the Inter American Statistical Institute the movement for regional organization is exploratory and tentative, with some reasons for belief, however, that it will spread and accelerate. The chief problems are geographic and organizational, the latter being the more difficult. In these respects, the administrative problem that confronts the United Nations is different from that involved in the statistical activities of the specialized agencies, where the major problems of coordination are those of subject matter competence and jurisdiction.

Two principal questions are raised by regional statistical organizations: What should be their chief functions in a world statistical organization; and what should be their relation to the United Nations on the one hand and to member countries on the other?

On the first point, it may be said that much of the interest in the further development of regional statistical organization centers on the benefit expected from the performance of various service functions, especially in the promotion of statistical education and the facilitation of common programs. Such service activities would not contradict but, indeed, would foster and develop the general statistical aims of the United Nations. Regional organizations so oriented in purpose and program would fit into world statistical organization without disharmony.

This then evokes the second principal question raised by regional

[5] The language of Article 52, Section 1 of the Charter is as follows: "Nothing in the present Charter precludes the existence of regional arrangements or agencies for dealing with such matters relating to the maintenance of international peace and security as are appropriate for regional action, provided that such arrangements or agencies and their activities are consistent with the Purposes and Principles of the United Nations."

statistical organizations. Some voice has been given to the view that the regional organizations might constitute an intermediate level between the United Nations and its affiliated specialized agencies and the member countries through which the international organizations would deal with member countries. There are forceful objections to this kind of arrangement. Excessive layers of authority between member nations and the central organs of the United Nations will tend to defeat that sense of participation in a world community which the United Nations Charter seeks to facilitate. It might happen that regional statistical organizations endowed with formal "line" authority would tend to displace the member nations as the basic elements making up a new world order. The losses in such an arrangement would fall not only upon the member nations deprived of direct relation with the central headquarters of the United Nations, but would prevent the latter from utilizing the full benefit of a strong central secretariat, the development of which would almost certainly be retarded if kept from direct contact with the principal producers of statistics—the member nations. The development and promotion of statistical education by the central statistical authorities would also tend to be handicapped if the latter had to function through intermediaries. Finally, the coverage of regional organizations is as yet confined to the Western Hemisphere. If regional statistical organizations were to be created universally to administer United Nations statistical programs, they would have to be created for that purpose by the United Nations. The strength of the IASI lies in its spontaneous and uncontrived evolution, out of the felt needs of the region it serves. Imposition of such an organization from above could create hazards of indifference or resistance which the United Nations cannot now afford, if indeed ever. The development of regional statistical organizations as service enterprises, however, rooted in the felt desires of the areas to be served, and devoted to statistical education in its many phases, promises interesting and significant developments for statistical science and administration.

PRIVATE INTERNATIONAL ORGANIZATIONS

Article 71 of the United Nations Charter empowers the Economic and Social Council to make suitable arrangements for consultation with non-governmental organizations on matters within its competence. The Statistical Commission in its capacity as adviser to the Economic and Social Council declared itself to be keenly aware of the important contributions to the improvement of world statistics which have been made by the International Statistical Institute and other organizations

in this field and expressed its desire that recognition be accorded to their work. It hoped particularly that appropriate means could be devised to bring the International Statistical Institute into harmonious and mutually advantageous relationship with the United Nations. One of the ways to do so would be to proceed under the authority of Article 71 of the Charter, although the Commission has deferred until a later time the formulation of recommendations as to specific methods by which such recognition might be expressed through the United Nations, and as to ways in which such organizations might be related to the United Nations and their activities utilized in fostering international cooperation in the improvement of statistics. The knitting of various private international statistical organizations into the world system is, therefore, a concern of the United Nations.

The Statistical Commission directed attention to the Institute (sometimes called ISI) because it is one of the oldest scientific organizations and because of its commanding reputation and influence in international statistical affairs. The ISI assembled for its last world conference in Prague in September 1938 but was forced by the imminent Nazi invasion to abandon its meeting in a hastily called midnight parley, without doing any business. Its last regular session was held therefore a decade ago in Athens; since then there has been no new election of officers nor other official business except that conducted through the Permanent Office at the Hague. Even this was interrupted by the war and the occupation of the Netherlands, which disrupted communications with the outside world. Although the Institute is privately organized and administered, having come into existence as the successor to the international statistical congresses initiated by the Belgian Quetelet, it was and is supported in part by subventions from the principal governments of the world, including the United States. The usual United States contribution was last received for the fiscal year 1939-1940. During the war, the appropriation was omitted from the budget estimates of the Department of State, but it has been restored in the estimates for fiscal year 1947. Plans and arrangements are being made for the Twenty-fifth Session of the Institute to be held in Washington September 13–25, 1947. There is general agreement that this session will be of the utmost importance, not only for the Institute but for the development of international statistical activities everywhere within the new world order. The Institute's own organization must be repaired and revitalized after the attrition of ten years, and its future course must be redrawn and redirected.

In the course of time the Permanent Office has tended to concen-

trate its attention upon demographic problems and the suggestion has been made that the United Nations might use the Permanent Office of the Institute for whatever work in demography the United Nations may require. Consideration of this course, however, will have to take account of the plans being made to establish a Demographic Commission within the structure of the United Nations.[6] Even if these plans did not exist, it would probably be preferable if the collection and assembly of international statistical data were centralized in the formal institutions of the United Nations and its affiliated agencies. To separate and parcel out various subject matter fields of statistics would introduce an element of dispersion contrary to the synthesis which is forming around the United Nations.

There are other and more feasible suggestions for the future role of the ISI in a world statistical order. Some members of the Statistical Commission have thought it might assume leadership in the development of world statistical systems. This is analogous to the activities of the Inter American Statistical Institute in the western hemisphere, which have proved very useful. The main techniques would presumably be those of advice, education, and promotion in statistically undeveloped areas of the world. Another proposal is that the Institute maintain a world statistical research center at the site of its Permanent Office, a place at which the frontiers of statistical research would be explored. Under this proposal, arrangements could be made to provide world statistical scholars with leaves of absence to serve as directors of research in residence, for fixed terms. Such a role would provide the Permanent Office with a program of continuing responsibilities, shifting its activities from those of a casual and episodic type, to those permitting long range planning and development. Still another proposal for the Institute would utilize it as a "supreme court" of statistics, an international academy, to which various questions of professional competence would be referred for judgment and opinion by the United Nations and member nations, very much as the American Statistical Association is occasionally called upon by agencies of the United States Government to study and make recommendations upon statistical questions. Finally,

[6] The United Nations Preparatory Commission recommended to the Economic and Social Council that the latter should consider "the desirability of establishing at an early date, and possibly at its first session" three commissions of which one was to be a Demographic Commission. The functions suggested for the proposed Demographic Commission were those of study and advice to the Council on matters related to: (1) population growth and the factors determining it; (2) the effectiveness of policies designed to influence these factors; (3) the bearing of population changes on economic and social conditions; and (4) general population and immigration questions. The matter was referred to the Social Committee of the Economic and Social Council (a drafting committee) which has not yet made its recommendations. See "Report of the Preparatory Commission of the United Nations," 1945, pp. 28, 38.

it has been suggested that the Institute might perform services for the improvement of the work and methods of the statistical fraternity much as the American Bar Association looks after the interests of the legal craft and seeks to improve its professional standing and practices.

There may be various other possibilities and the proposals made are hardly more than suggestions of alternative lines of development that have been mentioned by those interested in the problem that confronts the Institute. As an independent private group, the Institute will itself have to decide upon the role it wants to follow in the future; the United Nations will not dictate the course it takes. The United Nations, however, would welcome the affiliation of the Institute in some useful and practicable permanent relationship; and there seems to be ample room for a negotiation of terms satisfactory to both institutions. A meeting of the Statistical Commission will be held at about the same time as the scheduled Twenty-fifth Session of the Institute in Washington. By that time, specific proposals of agreement may have been worked out for consideration by both the Commission and the Institute, and it may then be possible to effect an agreeable arrangement fixing the Institute's permanent role and future relationship in the world statistical system. The extent to which other international statistical organizations are brought within the conspectus of the United Nations is a question upon which it is difficult yet to make any judgment; but the doors of the United Nations have been opened wide by the Statistical Commission.

NATIONAL STATISTICAL SYSTEMS

Action parallel to the formation of an international statistical system will be required in many countries, to order arrangements in such a way that maximum participation in international statistical activities will result. In countries which are statistically undeveloped, the organization of a basic national statistical system will be necessary, perhaps with the help and advice of the statistical services of the United Nations, its organs, or affiliates. Even in statistically developed countries, it will probably be necessary to designate focal points for dealing with the international organizations. In countries whose statistical services are highly centralized, the focal point ordinarily will be the central statistical bureau. In countries with decentralized statistical services, some focal point will have to be organized. In the United States, the statistical agencies operate under conditions of decentralization; it will be necessary therefore to create a clearing house for

dealings with international statistical agencies and steps to do so are going forward.

The first step was taken over a year ago when the conference of Federal statistical officials was called to consider the preparation of a statement of the need for international statistical services for use by the United States delegation at the San Francisco Conference. More recently, further steps have been taken to make a permanent organization out of the conference. On March 21 of this year, the same group of officials met at the invitation of the Division of Statistical Standards of the Bureau of the Budget to organize themselves as the Federal Committee on International Statistics. The membership of this committee is drawn from the principal agencies having important international statistical interests, including State, Treasury, Agriculture, Commerce, Labor, Export-Import Bank, Securities and Exchange Commisssion, United States Public Health Service, Office of Education, United States Maritime Commission, United States Tariff Commission, Federal Reserve System, Social Security Board, and the Civil Aeronautics Board. One of its first undertakings was the preparation of an inventory of international statistical activities carried on by the Federal Government.

The chief concerns of the Federal Committee on International Statistics will be the clearance of demand and supply arrangements with international bodies for the receipt and supply of data; the coordination of the international statistical requirements of the United States Government; the avoidance of duplication in the collection of statistics required by the United Nations, its organs and affiliates; and the clearance of intern arrangements and the loan of experts. Within the framework of coordination by the Committee, there will be every encouragement for direct contacts between the statistical technicians and specialists in the United States and their colleagues in other parts of the world. Full utilization of the statistical resources of the United States will be fostered through direct dealings with international organizations, with the knowledge and guidance of the Federal Committee on International Statistics.

Three specific instances of the way in which the Federal Committee will operate can be cited. First, the recommendation of the Economic Employment Commission already mentioned will, if it is approved by the Economic and Social Council, require the assistance of several Federal agencies in the collection or assembly of complete data on plant capacity and labor force. How should this assignment be distributed

among the Federal statistical family? It would seem evident that some central action may be necessary, and this action the Federal Committee would be in a position to supply; although it would not dictate the distribution, the distribution would be made with its assistance and certainly with its knowledge. Second, the inventory of Federal international statistical activities which is being made at the suggestion of the Federal Committee will lay the foundation for the designation of focal points among the Federal agencies for dealing with certain kinds of international statistical business. It will also make it possible to estimate the amount, kind, and incidence of international statistical business and the responsibilities which participation in the world of statistical system imposes. Third, the Federal Committee will be available as a starting and reporting point for missions of United States officials abroad on statistical business. In this fashion, traveling officials will have the benefit of the advice and counsel of the Federal statistical family, will be able to represent the views of that group on proper occasions, and will be able to perform various kinds of service abroad of benefit to American statisticians, such as getting in touch with those whose activities and whereabouts are in doubt.

These, at least, are the plans. The agencies of the United States Government have a stake in the success of world statistical organization. Although the committee device is no sovereign cure for the risks of administrative dispersion and diffusion, the organization of the Federal Committee on International Statistics may help the attainment of statistical unity in international affairs by promoting it at home. Because the Commission is sponsored by the Division of Statistical Standards of the United States Bureau of the Budget, its activities become part of the general program for the coordination and improvement of statistics among all Federal agencies, for which the Division has been made responsible. Under these auspices, the statistical world abroad joins the statistical world at home in common purpose and program.

OBJECTIVES, USES AND TYPES OF LABOR FORCE DATA IN RELATION TO ECONOMIC POLICY*

LOUIS J. DUCOFF AND MARGARET JARMAN HAGOOD
Bureau of Agricultural Economics

The kinds of labor force statistics developed should be determined by their uses. The full employment goal of national economic policy in the United States gives rise to the need for labor force statistics which can serve as a barometer of the state of functioning of the economy and which can identify the sectors of the Nation's workers for whom the economy is not providing full employment opportunity. This treatment is primarily of the implications of such objectives for the types of labor force data needed, although it is recognized that labor force data have many other uses.

Labor force statistics include data obtained from employing establishments, registrations, and population surveys, with each type having advantage for certain types of uses. Labor force data from population surveys need to be expanded to provide differentiated categories of the unemployed and to identify those employed workers whose employment is inadequate because it is insufficient in amount or is remunerated at substandard rates. The development of these and further geographic differentiations would increase the utility of labor force data in relation to economic policy.

THE USES to be made of any statistical series have important implications for the types and nature of data which should be obtained. The uses determine the sources, concepts, differentiations, and frequency of issuance of counts or estimates of various types of economic phenomena. Because the labor market activity of persons of working age affects every sector of the economy, labor force statistics have a wide variety of uses. In determining the concepts and differentiations that underlie the labor force data currently obtained, many types of uses need to be taken into account.

In this paper we will indicate some major objectives of labor force data and the ways in which different types of labor force data serve these objectives. We will then discuss underlying concepts and certain developments which need to occur to broaden the uses of labor force statistics. In covering these subjects, the full employment goal of national economic policy in the United States will be singled out for special emphasis, although it should be recognized that the needed de-

* This paper and the papers by Gertrude Bancroft and Emmett H. Welch, and Charles Stewart and Loring Wood, which follow it immediately, were planned as a unit and were presented at the session on Labor Force Measurement and National Employment Policy at the 105th Annual Meeting of the American Statistical Association at Cleveland, Ohio, January 24, 1946.

velopments in labor force statistics are not at all dependent upon the inauguration of any full employment policies.

I. OBJECTIVES OF LABOR FORCE, EMPLOYMENT AND UNEMPLOYMENT DATA IN RELATION TO ECONOMIC POLICY

A major objective of labor force statistics is to serve as a barometer which indicates currently the state of functioning of the economy as this is manifested in levels of employment and unemployment of the Nation's workers. Other types of statistics on production, income, etc., also indicate the general state of the economy, but no other type reflects as directly the changes in work opportunities. Labor force statistics are obviously required in planning or appraising *general* economic and labor policy, as well as *specific* measures designed to implement such policy. In this function these statistics serve not only government agencies, but also management, labor, and various groups with diverse interests in economic and labor policy.

Another objective of labor force statistics is to aid in diagnosing a given national economic situation through identification of the sectors of the Nation's workers for whom the economy is not providing full employment opportunity. In times of less than full employment, it is not enough to know simply the total number of unemployed workers. To diagnose the situation in order to appraise the steps proposed for remedying it, much more detail is needed. The location and characteristics of the wholly unemployed are needed, as well as their aggregate volume. Among those who have some employment, identification is needed of *underemployed* workers—that is, the partially unemployed. There is also the need to identify among the employed those who are remunerated at rates below some minimum level of adequacy implied in the goal or standard of "full employment opportunity."

Changes in the level of functioning of the economy also should be diagnosed according to the industries which are expanding or curtailing employment. Employment statistics may be viewed both as a measure of opportunities afforded workers, and as a measure of the state of functioning of various industrial sectors of the economy. Employment data by industries are especially needed for the planning of specific policies and programs relating to particular industries.

In the formulation or appraisal of economic policy labor force statistics serve another objective. They supply the most important part of the quantitative basis for making labor force projections into the future, under various assumptions as to the state of functioning of the economy. The projections made under the assumption of full employment have especial usefulness. On the one hand, they can provide a

more explicit formulation of what the goal of peacetime full employment would mean in terms of number of jobs to be provided. On the other hand, full employment projections can serve as a standard against which a given current national economic situation can be diagnostically appraised.

In summary, therefore, labor force data must be designed to indicate the state of functioning of the economy in providing employment opportunities, to identify and measure the groups of workers for whom full employment has not been obtained, and to help formulate national employment goals.

II. TYPES OF LABOR FORCE, EMPLOYMENT AND UNEMPLOYMENT DATA REQUIRED FOR VARIOUS USES

Sources of employment and unemployment data. The term "labor force data" in a broad sense includes all types of data, however derived, on employment and unemployment or the total labor force. In technical usage, the term "labor force data" is restricted to data from only one of the three principal sources of data on employment and unemployment. In this narrower sense, statistics developed from classification of the population of working age according to their current labor market status are designated as "labor force statistics," or sometimes as population classification statistics. Determination of the current labor market status of individuals provides the basis for estimates or counts of the total labor force, its employed and unemployed components, and of major categories of the remaining nonworker population. Labor force statistics of this type were obtained in the 1940 Population Census for the entire population of working age, and have been collected on a sample basis since then in surveys known as the Monthly Report on the Labor Force, initiated by WPA under the direction of Howard B. Myers, and transferred to the Bureau of the Census in 1942.

A second source of current employment data is the reports of establishments employing workers. Reports may be made on a voluntary basis, as in the case of nonagricultural establishments which report the number of their employees to the Bureau of Labor Statistics, and of farmers known as crop reporters who report their farm employment to the Bureau of Agricultural Economics. Reports may be legally required from establishments meeting certain criteria, as in the case of reporting required from employers in the administration of Old Age and Survivor's Insurance and Unemployment Compensation programs.[1] The biennial Censuses of Business and of Manufactures which were interrupted

[1] The importance of the difference between voluntary and legally required reporting rests on the fact that the former is subject to response bias while the latter is not.

during the war also obtained employment data through reports from employers. Employment data secured in these ways are known as establishment reports as distinguished from the type of employment data obtained from population classification in labor force surveys.

In addition to population surveys, another source of unemployment data is the registration of individuals who file claims for unemployment compensation at local offices of each State's Unemployment Compensation Agency. Compilation of these registrations permits a national total for workers who are claiming these benefits.

Uses served by the various types. Each of these three types of data can serve important uses. Labor force data derived from population surveys have the following advantages which make them superior for certain uses.

1. They provide the only source for estimates of the total labor force, for total employment and for total unemployment, without duplication between industries within the employed group, or between the employed and unemployed classification. Thus for a current cross-section view of utilization of the Nation's manpower, or for the record of change in the total labor force and its components, labor force data derived from population surveys provide the only complete picture. For these uses, it is essential that the estimates of the various components of the labor force be additive.
2. They provide the only source for deriving labor force participation rates—that is, the proportion of each age-sex group who are engaged in labor market activity. Labor force participation rates are used in current analyses and are the starting point in any full-employment projection for obtaining estimates of the number of workers for whom jobs must be provided at some given future date. Labor force statistics obtained from a population classification are therefore essential for the setting of full-employment goals.
3. Labor force data provide the possibility of subclassifications within the employed, unemployed and nonworker categories which can usually be made only when the report is obtained on an individual basis rather than from an employer reporting for a group of workers. Some of the important items now secured for further differentiation are age, sex, veterans' status, industry, occupation and time worked during the reporting week. Other items are proposed later for further differentiation of the employed and unemployed which would enhance the diagnostic function of these statistics.

On the other hand, establishment reported statistics on employment

have certain advantages over those derived from population surveys, which make them superior for certain uses.

1. Because they are obtained for relatively large groups of workers in any one report and usually by mail, establishment reported statistics are less expensive than those obtained from population enumeration. Hence, they can be obtained for finer geographic and industrial breaks without incurring prohibitive costs. Although current employment estimates for regions, larger states and the more important metropolitan areas could be obtained through the expansion of current labor force surveys proposed by the Bureau of the Census, establishment-reported statistics on employment are now the only type available in some geographic detail, except where surveys have been conducted for individual cities. In general, therefore, analyses of current employment requiring geographic detail now necessarily have to be based on establishment-reported data.
2. Establishment-reported data provide superior industrial classifications of the employed because: (a) the reports on employment are accompanied by other information which permits an accurate classification of the establishment by standard industry groupings, (b) the system of collection is by industry which makes for better representation of industries than in the case of population surveys in which other criteria are the major basis of stratification, (c) because of the first major advantage mentioned, namely inexpensiveness, the separate industries are more adequately covered than in population surveys. Therefore, establishment reported statistics have definite advantages in employment analyses for specific industries.
3. Establishments reporting employment data can and often do supply related information on production, man-hours worked, and wages which permit productivity and wage analyses by industries.[2] Although data on wage earnings have been occasionally secured in population surveys, those reported by employers are doubtless more accurate generally, since their payrolls are a matter of record. However, they have limitations for certain purposes since they do not represent total earnings of individual workers in many cases.

The unemployment data derived from registration of claims for unemployment compensation have present and potential uses. At present,

[2] Man-hours data for a given industry as reported by establishments are more accurate because in population surveys the total time worked during the reporting period by individuals who hold more than one job is all allotted to the one industry in which greater time was worked.

their usefulness is limited because not all types of employment are covered by the legislation, because States vary in their requirements for eligibility, and because compilations at the national level are therefore somewhat noncomparable and difficult to interpret. On the other hand, if these difficulties were overcome, there would become available unemployment data with a very considerable amount of geographic detail.

Reference is made in the last paper on this program* to specific examples of the uses of population and establishment reported statistics. The remainder of this paper is devoted to consideration of what the uses of labor force statistics in relation to economic policy imply for the concepts and classifications to be used in labor force surveys involving population classification.

III. IMPLICATIONS OF USES FOR BASIC CONCEPTS IN LABOR FORCE SURVEYS

One implication of the uses of employment, unemployment and labor force data in relation to economic policy is that the basic concepts must be designed for reflecting *changes* in the national employment scene. This means that the concepts of employment and unemployment must relate to a relatively short time period and that the measurement process must be made at frequent intervals. Both the 1940 Population Census and the MRLF surveys afford labor force statistics geared to activity or status during a specified current week. The use of labor force statistics to indicate short-time as well as long-time changes requires that the basic concepts of employment and unemployment should relate to a specified short-time period rather than to a "usual" status, such as in the "gainful worker" figures of the decennial censuses prior to 1940.[3] It is recognized, however, that for certain purposes, classifications based on longer time periods may be preferable, or at least desirable for supplementation.

Another implication of the uses of labor force data for the concepts is that they must provide for a classification of *all* persons of working age into labor force status groups which are mutually exclusive, so that the resulting estimates will be additive. This requires carefully specified priorities of status for those persons who have dual status during the reporting week. The priorities of status in current labor force measurement are: (1) at work, (2) unemployed, (3) with a job but not at work, and (4) nonworker status. Those at work and those with a job but not

* See page 313.

[3] Since the monthly labor force surveys of the Bureau of the Census are the only source of current national statistics derived from labor force classification of individuals, the rest of this paper refers primarily to the labor force statistics developed from the MRLF surveys. However, many of the matters discussed are generic to any labor force statistics, whether gathered for a city or some other area, as well as for the country as a whole, and whether gathered monthly, yearly, or in a one-time survey.

at work are added to get the total employed. In dual status cases, the person is classified in the category with higher priority. For example, persons who were at work for some part of the week, but unemployed and seeking work during an even larger part of the week are classified as employed. If a person spent most of the week in a nonworker status such as doing housework at home or going to school, but did gainful work during some part of the week, he or she is also classified as employed. A person whose major time is spent in some nonworker activity, or a person who has a job at which he did not work but who reported looking for work during the week is classified as unemployed.

The uses of labor force data required that some set of priorities be adopted which could be precisely defined and put into effect with a fair degree of uniformity. They require further that borderline groups be expressly allotted to one status or another so that they too will be treated uniformly. Consideration is being given to the possibility of changing the allocation of certain borderline groups. Certain groups of workers now classified as "with a job but not at work" are being considered for inclusion among the "unemployed," which has a higher priority. Persons on lay-offs of less than 30 days and certain other numerically small groups who did not report looking for work have been classified as "with a job" and included among the employed. Because they were "involuntarily idle," however, consideration is being given to including them with the unemployed. On the other hand, persons without a job who report that they were not looking for work because of temporary sickness or because they believed no work available are classified as unemployed, although the case could be made that some of these should be treated as nonworkers.

To reflect accurately changes in employment and unemployment, the criteria must be adapted to distinguish the employed, the unemployed and nonworkers in periods of varying economic conditions. Because different national situations call forth different types and degrees of labor market activity on the part of individuals, the general principle observed in the development of labor force concepts has been toward making the "in labor force" category as inclusive as possible. This calls for introducing further differentiations according to the degree of participation so as to permit measuring changes for comparable groups.

IV. IMPLICATIONS OF USES FOR DIFFERENTIATIONS WITHIN THE EMPLOYED AND UNEMPLOYED COMPONENTS OF THE LABOR FORCE

The barometric, diagnostic, and projection uses of labor force statistics have gained importance as the idea of full employment became

more widely accepted as a goal for the Nation's peacetime economy. Moreover, the growth of the idea of full employment has meant that further differentiations of the employed and unemployed are required to serve the uses described.

The virtual elimination of unemployment during wartime raised the challenge to eliminate all but frictional unemployment in peacetime, and "full employment" was defined as a situation with unemployment at a minimum or frictional level. But this naturally led to a complementary definition in terms of the number of jobs needed to keep unemployment to the specified minimum level. Projecting the number of jobs required for full employment rests primarily on projecting the number of persons who will be in the labor force. Such projections take as a starting point the past record of labor force participation of the different age-sex groups of the population under varying economic conditions. Thus differentiations of labor force statistics by age, sex, marital status, etc., are needed as a primary basis for the labor force projections which make explicit what full employment goals mean in terms of the number of workers to be supplied with jobs.

The growth of the idea of peacetime full employment did not stop with the mere specification of a sufficient *number* of jobs. More recently, various writers and groups have added criteria as to the *kinds* of jobs consistent with the idea of full employment. Full employment as a goal means, at least for some, a sufficient number of productive jobs, adequate in the regularity and amount of work afforded, and providing adequate remuneration.

Such an expanded goal sets new problems in determining and developing the differentiations within the employed and unemployed which would best serve the function of indicating the extent and nature of departure from the goal which a given national employment situation represents. The problems of full employment projections also require certain types of labor force data which have not been previously available. Many of the needs could be met by further classifications of the employed and unemployed according to data now being obtained. Other needs would require additional information which could be obtained in current labor force surveys, particularly if the size of the sample were somewhat increased.

Perhaps the most important additional differentiation needed in current labor force statistics is geographic. Especially if labor force statistics are to serve a diagnostic function, they must indicate *where* the trouble spots are. Complete geographic identification of the individuals covered by the MRLF surveys is already obtained on the schedule. But an expansion of the scope of the survey operations would be required to provide samples large enough to permit valid estimates of

current employment and unemployment for regions, by residence groups of the population, for metropolitan cities, and for States.

Differentiations needed among the employed. The most important need for additional differentiation among the employed is to separate those workers whose employment is sufficient and remunerative from those whose employment fails to meet these criteria. In a fuller treatment of this problem, we have for want of accepted terminology referred to the first group as "adequately employed" and to the second group as "inadequately employed."[4] Measurement of the inadequately employed may be as important as measurement of the unemployed for indicating the degree to which a given national situation falls short of meeting full employment goals.

Measurement of the inadequately employed would involve the development of labor force survey techniques for identifying among the employed: (a) those who do not have a sufficient *amount* of work, the underemployed, and (b) those who get substandard returns per hour of work because of its low productivity (mainly self-employed or unpaid family workers) or because they are working for substandard wages. In appraising the state of functioning of the economy at any given time, the size of these two groups as well as the number of wholly unemployed needs to be known to indicate the total number of workers for whom full employment has not been attained. In times of depressed economic conditions especially, the presence of large numbers of underemployed among those classified as employed would need to be recognized and estimated separately from the voluntarily part-time workers. Even more continuously, in times of high as well as low employment levels, is there need for measuring the number of workers who have a sufficient amount of work but who are inadequately remunerated. Especially as the nation approaches peacetime full employment goals in terms of *number* of jobs will attention need to be directed toward providing employment data which identify the groups of workers whose jobs or enterprises do not meet minimum standards of adequacy.

Differentiations needed among the unemployed. During the past, the total number of unemployed persons has been relied upon as an important indicator of the state of functioning of the economy. In the present transition period from war to peace unemployment may be of relatively short duration for most of the workers involved. A substantial proportion of the unemployed may be extra workers who came into the labor force in response to the economic situation brought about by war and who are in a state of indecision as to whether to

[4] "Labor Force Definitions and Measurement in Relation to Employment and Income Levels," preliminary draft of a report prepared for the Subcommittee on Labor Statistics of the Labor Market Research Committee of the Social Science Research Council, November 1945.

continue gainful work or to withdraw from the labor force. A few years later, the total number of unemployed workers may be of a very different composition. To interpret the meaning of a given national level of unemployment, it will be necessary to differentiate the total according to the degree of the worker's attachment to the labor force and duration of unemployment, in addition to certain subclassifications such as age, sex, marital status, residence and occupational skills.

The following four classes of the unemployed are suggested as the types of differentiations needed in using unemployment data in diagnosing a given national employment situation:

1. Unemployed persons who are customarily members of the labor force and who are seeking full-time jobs;
2. Unemployed persons who are customarily members of the labor force but who are seeking only part-time employment;
3. Unemployed persons who have never had jobs but who are seeking to enter the labor force on a permanent basis;
4. Unemployed persons who have not customarily been in the labor force and who are seeking only temporary work because of some special need or because of unemployment of other breadwinners.

In addition to these classes, certain groups of the unemployed not readily identifiable need especial study. These include persons who would like to have regular work but who because of age, lack of occupational skill, physical or mental handicaps or discriminatory labor practices are considered unemployable.

Problems in obtaining differentiations needed. Some of the information which would be required for obtaining the differentiations needed of the employed and unemployed is already being obtained in the current MRLF surveys, but the size of the sample does not permit full utilization of it in cross-classifications. Obtaining other information would require special questions to be added to the schedule. Some of the questions would involve types of information differing from the type now obtained with respect to subject, for example, wages received during week—or with respect to time reference, for instance, major work status during last 12 months. While the data obtained should be as objective as possible, some questions of a more subjective nature than those now used might be required for identifying certain groups—for example, a question on desire for more work to identify the underemployed. Devising of schedule and survey techniques for achieving these differentiations poses challenging problems in the field of labor force measurement. The next paper will deal with some of these measurement problems.

RECENT EXPERIENCE WITH PROBLEMS OF LABOR FORCE MEASUREMENT

GERTRUDE BANCROFT AND EMMETT H. WELCH
Bureau of the Census

Different uses for labor force data often require different concepts and definitions which are sometimes difficult to reconcile. It is important to develop, as a part of measurement, procedures that will provide data which are suitable for the various uses and also comparable over time. The experience of the Census Bureau in the operation of its monthly sample survey during the last few years has revealed many of the problems attendant upon putting into actual practice the agreed upon concepts.

In a recurring enumeration of a sample of households, questions on which classifications are based must be such that they can be asked repeatedly. They must be simple and objective and not place too great a strain on the respondent's memory. *The basic classifications finally set up must be large enough* to be determined fairly reliable by the sample.

THE PRECEDING paper outlined some of the uses and purposes to be served by labor force data and indicated additional concepts and classifications of data needed. Recent experience in measuring and classifying the population in accordance with present labor force concepts indicates a number of problems that would have to be dealt with in expanding the pattern of concepts and classifications used. The first difficulty in obtaining adequate labor force measurements is to reach agreement on concepts and definitions of what is to be measured. Often different uses require different concepts and definitions. For example, if one is using unemployment figures to represent available labor supply, persons with a job but not working and not looking for work would not be included among the unemployed. On the other hand, if one is using unemployment figures to represent the interruption or lack of earnings, it might seem desirable to add to the unemployed those with a job but temporarily not working because of lay-off, strike, bad weather, etc. Present labor force measurements classify separately persons having jobs but not working for various reasons. It has been suggested that such persons might be classified as unemployed at one time and as employed at another, depending upon the purposes to be served by the data. This immediately introduces a second difficulty, namely, that it is not practicable to have two different unemployment series. A possible solution is not to have one unemployment series at all but to have a series representing persons looking for

work, another series representing persons with a job and not looking for work, etc. This solution would not please those persons who wish to have some simple, generally accepted term such as unemployment. The practice which has been followed up to the present and one which we personally believe may be found reasonably satisfactory for the future is to define unemployment in such a way as to be most generally useful and then to provide subclassifications of employment and unemployment that will enable other groupings to be prepared for special purposes. This approach seems desirable for another reason. It provides some flexibility in concept while providing continuity and comparability over time.

After a set of concepts and definitions acceptable to the users of the data has been obtained, another and more serious difficulty is to insure that the concepts and definitions that are agreed on are actually translated into measurement. The concept of employment in current use, for example, includes all persons engaged in some activity for pay or profit. This includes persons working for wages or salary, employers, own account workers, and persons working on a family farm or in a family business. It includes persons working long hours, those working short hours, and those with a job at which they are not currently working because of vacation, illness, bad weather, strike, or temporary lay-off.

Many persons have preconceived ideas of employment or unemployment which differ considerably from the concepts and definitions in terms of which measurements are being sought. The problem of obtaining an accurate measurement of labor force concepts by means of population enumeration and classification is to devise schedule questions and definitions which will insure that persons are classified in accordance with the desired concepts and definitions rather than in accordance with the respondent's ideas as to labor force classification.

It is to be expected that there will always be a considerable variation of response in the labor force classification of individuals, arising from difference in enumerator and difference in respondent. These variations in response become serious if groups of persons with certain characteristics tend to be persistently misclassified according to the concepts and definitions being applied. The experience we have had in measuring employment provides a rather striking illustration of this problem.

Since its inception, the Monthly Report on the Labor Force has attempted to achieve objective measurement of employment in terms of all persons who, during the census week, did any work for pay or profit, including those who worked on a family farm or business without

specific wages. During the past several years considerable evidence has accumulated to indicate that the enumeration of employment in the 1940 Census of Population and in the Monthly Report on the Labor Force, or the MRLF, had been incomplete because of: first, a failure to include as employed a considerable number of persons such as housewives and students who did not consider themselves primarily as workers and second, a failure to include large numbers of unpaid family workers.

Under-enumeration of employment by the 1940 Census of Population and by the MRLF is indicated by a comparison of employment figures obtained in the Census of Population and in the MRLF with those obtained through establishment reports by the Census of Agriculture, the Bureau of Labor Statistics, the Bureau of Agricultural Economics and the Social Security Board. Estimates of employment based on establishment reports would always be expected to exceed estimates based on population enumeration because individuals working in more than one establishment during a reporting period are counted more than once. However, the extent to which the employment estimates based on establishment reports have exceeded the estimates based on population enumeration has been greater than the expected duplication in establishment reports.

Enumeration experience also indicated that an incomplete employment count was being obtained. In March, 1942 each person who was neither working nor looking for work was asked if he could take a full-time job if one became available within 30 days. This inquiry so influenced the responses to the regular labor force questions that the estimated level of employment increased by almost a million, and the number of persons classified as housewives and students decreased by a corresponding amount. In the following months when the MRLF interviews were conducted as usual this increase in the civilian labor force vanished. In November, 1942 a second inquiry was made into the available labor reserve with similar results.

Repeated evidence of this type, even after careful training of enumerators in handling supplemental questions, suggested that the additional questions did not distort the MRLF results, but rather that they partially corrected for a consistent under-enumeration of employment. As a result of those experiences, it was decided to make a more systematic investigation of the nature and extent of the under-enumeration of employment which was occurring.

The evidence at hand suggested that the question, "Was this person at work at a private or government job last week?" does not ob-

tain a positive response for some employed persons who consider themselves to be not workers but housewives or students. Evidence available also indicated that a large proportion of unpaid family workers were not counted as employed because the wording on the schedule suggests paid employment. In addition, the exclusion of incidental chores from the definition of unpaid family work operated to exclude some persons working substantial amounts of time.

Various checks indicated that there was little uniformity from enumerator to enumerator or from respondent to respondent in what is considered to be incidental chores. It appeared that more consistent and reliable results would probably be produced by including chores as unpaid family work and then eliminating from the count of unpaid family workers those working fewer than a specified number of hours. (It is necessary to attempt some exclusion of incidental chores from unpaid family work; otherwise an employment count among persons living on farms or operating family business approaches a count of the population of working age.)

In the fall of 1944, the Census staff began developing and testing a revision of the MRLF schedule with a view to achieving a more nearly complete count of all employed persons. By January 1945, we had a revised schedule and definitions which sought to correct the difficulties in the old schedule as follows:

The first question on the new schedule merely asks what the person's major activity was during the census week. Most people tend to think of themselves as engaged primarily in some one activity, working, keeping house, going to school, etc., even though they may also be engaged in various other pursuits. The enumerator using the new schedule accepts the respondent's statement of his major activity, knowing that some persons in the labor force will not report themselves as workers at this point in the interview. Then for all persons whose major activity or status is indicated as something other than working, the enumerator asks whether, in addition to his stated major activity, the person being enumerated did any work for pay or profit during the census week. It is at this point that the part-time work of the housewife or student is naturally reported. Note that the old schedule attempted to obtain a complete count of persons at work by asking the single question, "Was this person at work on a private or government job last week?" The new schedule, on the other hand, uses two questions. The first enables the respondent to give his own classification of activity status and the second asks specifically whether or not persons

whose major activity is considered to be something other than working, did any work during the census week. Many persons were surprised at the effect that this change in the method of obtaining a count of employment had upon the number of persons reported to be working. The question on the old schledule appeared to be quite definite in its meaning. Actually, as we have seen, it did not mean the same thing to different people.

The new schedule definition of unpaid family work includes incidental chores. As a substitute for the exclusion of incidental chores in the old definition, persons working less than 15 hours per week at unpaid family work are excluded in the tabulation process from the count of those at work.

The new schedule was pretested in April, 1945, in all MRLF sample areas throughout the country. The pretest was based on a sample of approximately 2,000 households selected at random from the total MRLF sample. The sample households were enumerated with the old schedule as a part of the regular April enumeration. The following week the sample households were enumerated a second time, using the new schedule. The information recorded on the new schedule applied to the same census week as the old schedule. The employment status information from the old schedule together with the information from the new schedule was transcribed to punch cards and tabulated so as to provide a direct comparison of the employment status information obtained by use of the two schedules. The tabulations indicated that the new schedule would increase the count of males employed by 900,000 and of females employed by 1,600,000. The count of unemployed among males was reduced by a little over 100,000 and was increased among females by about an equal amount. Over 90% of the additional male workers were under 20 years of age, most of them students. Among females, on the other hand, the additional workers were distributed among all age groups and most of them were previously classified as housewives. Of the additional workers found in nonagricultural industries over 50% were in trade and service activities, which are the fields of employment in which part-time work by housewives and students is most prevalent. While many of the additional workers worked only part time, the entire group could not be characterized as part-time workers. Nearly 50% of the additional workers worked 35 or more hours during the census week.

About one in ten of those classified as employed on the basis of the new schedule reported some other activity, such as home housework

or going to school, as their major activity. The old MRLF schedule recorded only about one-half of these persons as working and the other half as not working.

On the basis of the pretest results, it was decided to adopt the new schedule beginning in July 1945. In that month the entire sample was interviewed first with the old schedule and then with the new schedule to obtain a complete check of the difference in results obtained with the new as compared with the old schedule.

In July the differences in response obtained on the new schedule as compared with the old were similar to the differences obtained in the April pretest. The magnitudes, however, were somewhat different. In July the new schedule showed an increase of 1,400,000 in employment as compared with 2,500,000 in the April pretest. In large part the difference between the April and July results arose from the fact that in April school was in session and a large number of employed students were misclassified with the old schedule. In July, when schools were not in session, this occasion for misclassification did not exist. The April pretest also indicated a somewhat larger number of misclassified housewives than did the July double enumeration. This may have been due to the fact that some enumerators discovered during the period of discussion of the new schedule and the April pretest that the old schedule was obtaining an under-count of employment and they, therefore, adopted interviewing techniques that in part compensated for the under-enumeration of employment obtained with the old schedule. To reduce the extent of disruption of the MRLF series prior to the adoption of the new schedule, supervisors were instructed not to discuss the new schedule nor the reasons for its adoption with the enumerators; the April pretest was conducted with a small number of enumerators; and those who participated in the pretest were instructed not to discuss the results with other enumerators.

As we reported earlier, there is evidence indicating that under-reporting of employment existed in the 1940 Census of Population as well as in the MRLF and in other population enumerations which use the same type of questions to obtain a count of employment. The extent of under-reporting is much greater in those age and sex groups in which a smaller proportion are usually employed. These same age and sex groups have contributed most to the wartime expansion of the labor force. It seems probable, therefore, that the extent of under-reporting obtained with the old schedule was greater in 1944 and 1945 than in earlier years.

We are working on a revision of the MRLF data prior to July, 1945, to bring them in line with the measurements obtained with the new schedule after July, 1945. The procedure we are following is to relate the size of the adjustment to the level of employment for each age and sex group. This will result in a much less revision for in the earlier months than for the later months.

The new schedule has also affected the measurement of unemployment, but to a much lesser extent. Beginning in 1940 the basis for determining whether or not a person is unemployed was whether or not he was actually seeking a job. Persons who would be actively seeking work except for the fact that they were temporarily ill or that they believed there was no work available in the community, were also classified as unemployed and were referred to as inactive job seekers. The old schedule obtained a count of these persons by asking those who reported that they were not actively seeking work the question, "Why not?" In the months just prior to the adoption of the new schedule, when unemployment had been reduced to less than one million, nearly 50 per cent of the total number of unemployed were included in this inactive group. The new schedule does not ask the reason for not looking for work. Rather it relies on the respondent's answer to the question as to whether or not the person being enumerated is looking for work. If persons indicate in the course of the interview with the new schedule that they would be looking for work except that they are temporarily ill or believe no work available, they are classified as looking for work. The number of these inactives picked up with the new schedule, however, appears to be considerably less than the number counted with the old schedule which contained the question as to why persons were not actively seeking work. In addition, some of those reported as actively seeking work with the old schedule are found with the more specific and complete questions on the new schedule to be at work in addition to seeking another job. Such persons are classified as employed. Finally, the new schedule obtains a more nearly complete count of housewives and students who are actively seeking work. The net effect of all these factors on the count of unemployment obtained with the new schedule as compared with the old is small, the net change being a reduction of some 150,000 in the level of unemployment in July 1945.

Experience with the new schedule since the end of the war is providing some very encouraging evidence as to the sensitiveness of the looking for work question as a measurement of unemployment. For example, a large proportion of the women released from war plants im-

mediately started reporting themselves as not looking for work. Likewise a number of men who decided after the war to take a short rest or vacation from the labor force before again looking for a job were reported with the new schedule as not looking for work.

Our experience with measurement during the last few years emphasizes the difficulty of obtaining a full count of persons having a given characteristic—for example those who did any gainful work during a given week. Since such counts will include persons having a considerable range of characteristics, it is important to develop, as a part of measurement, a comprehensive and carefully worked out classification procedure that will provide the subdivisions of employment and unemployment that are needed for various uses and purposes.

At the same time, there are definite limitations on the refinement that can be introduced successfully into labor force classification based on a recurring enumeration of a sample of households. What sort of criteria have developed out of our experience?

First: The questions on which classifications are based must not arouse antagonism and must be such that they can be asked month after month. (In order to provide greater continuity in the data, MRLF households are enumerated monthly for about 6 months.) One of the reasons for the difficulty with the old schedule was that the enumerator was supposed to ask of all persons not working or looking for work, their reasons for not looking for work. Because this proved to be awkward in many cases, some enumerators probably did not ask the question, but classified the respondent by observation.

For the same reasons, the classification should not depend on a long series of questions which will discourage continued cooperation. Enumerators are only human and are likely to adopt short-cuts in order to avoid serious opposition, particularly when they have to return to the same household the next month.

Second: The questions must be objective and designed to provide approximately the same answers no matter who asks or answers them. In the great majority of cases, the information is furnished by the housewife. She can give fairly adequate replies to questions on whether the members of her family are working or looking for work, the approximate amount of time they worked and the type of industry in which they worked. She cannot be expected to give equally reliable answers to questions of the intentions, preferences, or attitudes of other members of the household. For example the question, "How many hours a week does your daughter want to work?" might be answered quite differently if asked of the daughter herself.

Third: The questions must not strain too much the respondent's memory. This criterion applies to the recency of the date or period of time to which the information requested applies, the length of the period of time to which the question applies, and the detail in which the information is requested.

Fourth: Concepts, definitions, and instructions must be simple and conform as nearly as possible to common usage. For example, we obtain a measurement of unemployment by simply asking whether a person is looking for work rather than by specifically asking whether or not a person was engaged in one or more of such activities as placing or answering ads, writing letters, applying at a factory personnel office, registering with the USES, etc. Of course, in instructions to enumerators it is necessary to indicate the criteria to be used if the respondent is not certain whether or not some member of the household is looking for work.

In setting up a labor force classification we must take into account not only what kind of questions will yield reliable and useful answers but also, in the case of sample surveys, what types of estimates can be provided by a sample of a given size.

Our fifth criterion, then, is that the classification should be sufficiently broad so that the month to month changes in estimates are significant and can be distinguished from sampling variation. For example, it has been proposed that unemployed married women whose husbands are earning $25 a week or more constitute an important group to break out from the total number looking for work, on the theory that their unemployment is not so serious as an indication of failure to achieve full employment. Even if the premise were agreed to, it would be a mistake to adopt such a classification if the sample cannot provide reliable estimates of the size of this group from month to month. Unwarranted conclusions would be drawn from changes which are only the result of sampling variation.

Again, differentiation is possible at certain times but not at others. The breakdown of the unemployed into the groups suggested in the previous paper might be proper when the level of unemployment is high but not when the level of unemployment is low. The groups in the population which it would seem desirable to differentiate because they do not have the characteristics generally associated with the majority tend to be small in size and therefore impossible to estimate adequately with a small sample.

These cautionary words do not mean that a wide variety of differentiation is not possible if obtained occasionally and on the basis of

special inquiries. Questions which would end all types of cooperation if asked each month can be asked once during the period in which a household is in the sample. Occasionally it would be possible to pay the cost of interviewing each adult in a household, so that he can report directly his present attachment to the labor force, his earnings if employed, and the type of work he is looking for if unemployed. Finally, the characteristics of very small segments of the population can be estimated from time to time, and although the estimates will be subject to a large sampling error, they will still be useful as an indication of the approximate size of the group, in relation to other groups and to the total.

The types of differentiation proposed in the preceding paper are, for the most part, consistent with the criteria proposed. Geographical differentiation and that related to such factors as age, sex, and marital status present no enumeration difficulties. The size of the present sample, however, will yield only national estimates. Funds are being requested to enable estimates to be made on a quarterly basis for 50 metropolitan areas and for the larger States. The expanded sample would also make it possible to provide much more extensive and reliable breakdowns of the national data.

EMPLOYMENT STATISTICS IN THE PLANNING OF A FULL-EMPLOYMENT PROGRAM

CHARLES STEWART
Bureau of Labor Statistics

AND

LORING WOOD
Bureau of the Budget

The term "full-employment program" usually means a program designed to maintain the general demand for labor. Unemployment can be reduced by such a program only to a certain point; beyond that point an increase in the general demand for labor will be relatively ineffective and will have undesirable repercussions. In setting a goal for a full-employment program we are essentially expressing a judgment on how far we can safely go in reducing unemployment by increasing general demand. In translating this judgment into figures, the best guide is our historical experience.

The essence of a full-employment program is the coordination of a multitude of governmental activities to achieve a quantitative result. This requires that the goal, initially expressed in employment terms, be translated into a comprehensive national budget. The construction of such a budget requires essentially the identification and projection of causally significant relationships. For this purpose, the primary statistical need is that economic data in various fields be made more comparable and consistent.

Objectives expressed in terms of national aggregates of employment, unemployment and expenditure are appropriate in the long-range planning of economic policy, but alone will not be sufficient for current appraisal of the adequacy of the measures taken. For this purpose we cannot rely on any single set of data. We must rather undertake a continuing analysis of all information which may throw light on the adequacy of our employment objectives and the meaning of current levels of employment and unemployment in the context of the developing situation.

RECENT discussion of full-employment policy places upon statisticians and economists a formidable task of developing the types of statistical analysis needed in its planning and execution. The task is not so simple as the discussion might seem to imply. The preceding papers have dealt with problems and recent progress in refining some of the statistical data which will be needed for such purposes. This paper deals more directly with some of the problems arising in connec-

tion with the use of labor force and employment statistics in the planning and administration of a full-employment program.

Three types of problems will be considered. First, there are problems of defining quantitatively the objective of full-employment policy, in terms of employment and unemployment. Second, there are problems centered around the development of the national budget as the central tool in economic policy-formation. Finally, there are problems connected with the use of statistical data in determining currently how closely the full-employment goal has been approached.

I

A common approach to the definition of full employment makes use of the relationship between the number of persons unemployed and the number of unfilled job openings. Beveridge, for example, suggests that full employment means "more vacant jobs than unemployed men."[1] To anyone concerned with problems of measurement, this type of definition raises serious questions. It is difficult enough to determine when a given individual is to be classified as unemployed, but there is at least no question that he is an individual, and, once his proper classification has been determined, he can be counted. A "job opening" however, is something much more tenuous; it may represent only the desire of an employer to interview prospective workers in order to be able to hire promptly at some future time in case the need should materialize.

We may get further if we ask why it is necessary, in setting up a "full employment" objective, to allow for any unemployment. It is doubtless true that unemployment of two million will involve somewhat less hardship to the individuals involved than will unemployment of 15 million. But this is not in itself a justification for taking two million unemployed, rather than one million or 100,000 as our "full employment" goal. The justification must be sought in the nature of the program for which the goal is being set.

The term "full-employment program", in the context of most current discussion, has a somewhat more specific connotation than the words themselves seem to imply. It ordinarily means a program designed primarily to maintain the general demand for labor, usually by maintaining the demand for goods and services. While some consideration may be given to demand for labor in specific regions and industries, emphasis is generally on the national picture. Measures designed specifically to improve the organization of the labor market—improvement of employment services, elimination of seasonal variations, or reor-

[1] Sir William H. Beveridge, *Full Employment in a Free Society*, New York, W. W. Norton and Co., 1945, p. 19

ganization of casual industries—are usually thought of in a separate category.

Increased demand for labor will undoubtedly result in *some* reduction in unemployment under any conditions. But if unemployment is already low, the effect will be relatively small, and serious inflationary pressure may result.

In setting a "goal" for a full-employment program, we must take account of the limitations of the program. The allowance for "minimum" or "frictional" unemployment which we make will vary, of course, depending on the character of the program itself and, in particular, on the extent to which it includes measures designed to improve the organization of the labor market. At best, however, the figure will represent a judgment on how far we can safely go in reducing unemployment through increasing demand.

In translating this judgement into figures, the best guide is probably our historical experience. At present, the only period of relatively full employment for which data are available is the war period, which is not an adequate guide to peacetime conditions. It is relevant, nevertheless, to consider that unemployment, as measured by the Census series, was below one million during most of the period from mid-1943 through mid-1945. We should probably not choose to face in peacetime the problems of labor shortage which characterized this period, and to invoke the manpower and price controls which they necessitated. The earlier part of the war period, before labor shortages became acute, may provide a better guide.

It is important to recognize that any projected unemployment or labor force figure has meaning only with reference to a specific set of definitions and techniques of measurement. The preceding papers have suggested the extent to which relatively minor changes in procedures of enumeration may affect the resulting estimates.

II

An objective defined only in terms of unemployment is hardly adequate as a guide to the development of economic policy. We need also to know what the goal implies in terms of employment, production, and national income. It may be necessary, indeed, to go much beyond this to the preparation of detailed estimates of income, consumption, savings, and investment, and perhaps even estimates of production and employment for individual industries.

It may not be obvious why such estimates are relevant to a program primarily concerned with employment. But the term "full employ-

ment" refers to the objective of the program, and not to its scope. What is implied by the phrase "full-employment program" is a coordination of Federal economic and fiscal policy in the interest of maintaining employment.

The essential characteristic of the program is thus the *coordination* of Federal activities in a multiplicity of fields which impinge on the operation of the economy. It involves, moreover, their coordination to achieve a *quantitative* result—a given level of demand for labor neither too high nor too low. The planning and execution of such a program will clearly require the preparation of comprehensive estimates covering the whole range of economic activity. The term "National Budget," frequently used to describe such estimates, is a fairly exact metaphor.

A full-employment budget starts with a forecast of the labor force, based on the latest current data and allowances for anticipated future developments. After deducting whatever unemployment figure is felt to be consistent with the full-employment objective, this yields a goal expressed in terms of total employment. From total employment we pass to estimates of production and income; and thence to an estimate of consumption, investment and total expenditures. The comparison between total expenditures and total production provides a test of the adequacy of demand to sustain full employment, and the underlying estimates provide, in greater or less detail, a description of the associated pattern of economic activity.

The problems involved in the construction of such a budget are in essence problems of identifying and projecting significant economic relationships. The task is complicated by the fact that most of the data which must be used were not designed for such purposes, and in many cases cannot be easily adapted to it. Two examples, both involving labor force and employment statistics, will serve to illustrate the sort of statistical problems which are encountered.

In passing from the estimate of total employment to estimates of production and income, we must make use of the projections of hours, man-hour productivity, and hourly or weekly earnings. Here we encounter the difficulty that our employment data are of two fundamentally different sorts, based on different concepts and definitions of employment. Data on labor force and unemployment are necessarily based on reports from the workers themselves (or members of their households) on the basis of which each individual is classified on the basis of his employment status. But such data do not lend themselves to the study of changes in productivity, hours, and earnings. For this

purpose, it is necessary to make use of reports from employers which provide data on the number of persons on their pay rolls, their aggregate earnings, and the aggregate number of man-hours worked. Such reports necessarily require the use of a somewhat different definition of employment, and the estimates of total employment derived from these reports may differ materially from those based on a household survey.

Lack of comparability between different series of data is not always traceable to such fundamental differences in techniques of measurement, however. The statistical work of the Federal Government was not designed to serve the needs of national economic policy-making. Rather it has developed gradually in response to much more particularized needs and interests. As a result, estimates of different but closely related things, prepared by different agencies for different purposes, may not be consistent with respect to scope, definitions, and classification even when the basic data are drawn from the same or similar sources.

Current data on hours and earnings in nonagricultural industries are available only from BLS reports, which also provide data on employment and aggregate wages. Trends in productivity, hours, and earnings must therefore be studied largely on the basis of these data. But to use the results of such studies in the construction of national budgets, they must be fitted into the structure of the Commerce Department estimates of national income. The estimates of the wage and salary component of national income, though based largely on the same sources as the BLS estimates of employment, are not directly comparable with the BLS estimates with respect to classification, or even with respect to total level and trend. Each agency has used the classifications and methods of estimation and adjustment it felt best adapted to its own needs, using the best data available at the time. This particular case is cited not because it implies criticism of either of the agencies involved; they have recognized the problem and are now seeking to achieve consistency. But it is illustrative of the problems involved in adapting to the needs of national economic policy-making a series of statistical programs that have developed largely in response to different needs.

III

The long-range planning of a full employment program may be carried on with primary reference to objectives expressed in terms of gross national product or expenditures. But as the expenditure goal is ap-

proached, it will be necessary to determine whether in fact the accompanying employment represents full employment, under-employment, or over employment. Presumably the objective would also be stated in terms of employment. But that goal likewise is tentative in character, and requires periodic checking to determine whether it actually represents, at any given time, the most accurate measure of the full-employment objective.

As and if we approach the stated goal, how do we know if the employment goal is actually consistent with the ultimate objective—the employment of the Nation's labor force? Until 1940 only decennial labor force data were available, which afforded scanty material for determining labor force norms. Estimating normal labor market participation rates and projecting actual labor force aggregates is still a virgin field of research. It is obvious, particularly at the present time, that forecasts of the probable size of the labor force for even one year in advance cannot be precise.

If we should fail by a wide margin to reach full employment, the facts of the situation would probably be clear, and more refined statistical analysis would hardly be necessary. But if we should approach full employment more closely, a more careful analysis would be required, and under some circumstances a reappraisal of the goal would be necessary. If, for example, the actual labor force, as shown by the Census Bureau's Monthly Report on the Labor Force, should prove to be substantially below the estimated labor force as used in determining the employment goal, downward revision would be required. On the other hand, if the estimates of required employment should prove to be too conservative, the continued use of too low a figure for the full employment objective would understate the real employment needs of the Nation and could result in a failure to take necessary action to utilize our real labor resources.

At this point we are primarily concerned with the problem of determining how we know whether or not we have full employment when we are at high levels of economic activity. We have to "get behind" the statistics and appraise their meaning in terms of the general labor market and economic situation. We need to provide for the continuous analysis of all relevant information which may throw light on the adequacy of the employment objective and the meaning of the prevailing levels of employment and unemployment in the context of the developing situations. More specifically, we need to examine periodically the reasonableness of the allowance made for frictional unemployment, the current level of the labor force as against projected levels, and

questions centering around the degree of under-employment of employed workers.

The volume of unemployment shown by the Census Bureau's Monthly Report on the Labor Force does not provide us with a conclusive answer as to whether the current employment level is actually approximating full employment when unemployment is relatively low. The primary reason for this ambiguity is the problem of appraising the volume of unavoidable frictional unemployment in any given situation. If this were readily determinable there would be strong reasons for adopting unemployment as the key measure—i.e., for using a stated minimum volume of unemployment as the objective. Probably the best clues as to whether unemployment at any given time is largely frictional can be found in evidences of labor shortages in some parts of the economy or in pressures toward rising prices for the factors of production. Hence over-all appraisal of the general economic situation must be relied upon to determine if the volume of unemployment reported at any given time is essentially frictional in character. It is apparent that the frictional component of unemployment cannot be readily isolated and may vary widely from time to time. Experience may prove that any rough allowance made in advance—for purposes of projecting an employment goal—may be so far off as to result in a serious distortion of the employment objective.

A second type of problem emerges from the possibility of over-statement of the size of the labor force as estimated currently on a household enumeration basis. The Census estimate of the number of persons seeking work but without jobs has the virtue of reflecting the respondents' own declarations that they are seeking work—a virtue which may give rise to the criticism that it is too subjective in character.

The general alternative to revising the employment goal to take account of the actual changes in the size of the labor force is to base the employment goal on long-run historical trends as to the proportions of the population of working age in the labor force. This may mean, in effect, that we hold to an arbitrary employment goal which has at best some valid relationship to labor market participation rates in 1940, and ignores marked changes which may have occurred since that time.

Any dangers inherent in the possibility of an overstatement of the numbers of persons reported currently as seeking work may be guarded against by continuous analysis of a wide variety of labor market information. An analysis of the age-sex composition of the labor force as reported in the Monthly Report on the Labor Force, for example, will reveal whether the number of youths, women, and older workers show

any unusual developments. It is possible, further, to examine the geographic and industrial composition of unemployment. The Monthly Report on the Labor Force does not provide the basis for such an analysis because it is not designed, at present, to show the geographic or industrial distribution of unemployment. But this can be approached by analysis of unemployment compensation claims data and analysis of changes in employment, by industry, by state, and by area, as shown by the Bureau of Labor Statistics reports, as well as by Employment Service data on job applicants and referrals.

The unemployment compensation data refer, of course, only to covered employment, but would reveal geographic concentrations as well as affording proof of availability and employability. Analysis of the BLS employment data, while not conclusive, would suggest whether excessive disemployment had occurred in any particular segment even if unemployment nationally was reasonably low. Employment Service data would reveal the age, sex, experience and other characteristics of job applicants as well as changes in job opportunities by area and industry.

Data of these kinds may throw little light on the question whether the prevailing volume of unemployment results from lack of adequate over-all demand for labor or from frictional causes. To the extent that the data point to regional or industrial concentration of unemployment, however, this would suggest that the unemployment arises from immobility of labor or other factors of production and from imperfect organization of the labor market. From the point of view of remedial action, this would indicate special measures rather than broad fiscal or other measures directed toward increasing aggregate demand. But in any case, such review and analysis of all available labor market information would provide a check on the validity of the current estimates of the size of the labor force as shown by the Monthly Report on the Labor Force.

A third problem, somewhat different in character, is the existence of conditions of under-employment—e.g., employment in marginal activities or at a workweek shorter than desired—as a result of lack of real full employment opportunities. With employment at relatively high levels, it may prove difficult to demonstrate the facts satisfactorily enough to justify remedial action. Information as to weekly hours of work are available from the Bureau of Labor Statistics monthly reports for many industries and from the Monthly Report on the Labor Force for a cross section of the population. Many individuals desire part-time employment, of course, and even where a marked increase

in such employment can be shown, the meaning is not altogether clear except in periods of recession or depression. Further research and information would be required, as indicated in a previous paper, for adequate evaluation of the existence of under-employment under conditions of relatively high employment. The problem is significant, however, because the existence of under-employment presents evidence as to lack of over-all demand for labor as a factor impeding the most efficient utilization of the Nation's labor resources.

MEASURING AND FORECASTING CONSUMPTION*

FRANK R. GARFIELD

During the past decade, under the stimulus of changed economic conditions and changed attitudes toward policy formation, real progress has been made in the measurement of consumption in the United States. Concepts have been clarified considerably, techniques for gathering information have been improved, some information has been collected, and a certain amount of lay as well as professional support has been developed for further use of public funds to measure consumption. Work in this field, however, is still in a pioneering stage, particularly with reference to the measurement of changes in the physical volume of consumption.

One purpose for which information on consumption has been used increasingly during recent years is the forecasting of changes in consumption and in economic conditions generally. Consumption has been widely regarded as a variable dependent in quite regular fashion on certain other variables, particularly consumer income after taxes and increments therein. This interpretation has facilitated calculations of hypothetical figures for future gross national product and for future employment; and has contributed also to conclusions reached about prospective aggregate demand and aggregate supply, and prospects for inflation.

In the author's view, generalizations about the relationships between consumption and other elements in the economic situation have been made without adequate recognition of differences in behavior resulting from differences in time, place, and circumstance. This accounts in part for the shortcomings of many forecasts made last summer. Five suggestions concerning forecasting are offered for consideration: (1) that consumption in any particular period be estimated after close study of the prospects for component parts, with income after taxes being regarded as only one (very important) factor influencing consumption; (2) that more attention be given to accurate measurement of the current level of the physical volume of consumption as a starting point for estimating future changes in consumption; (3) that the common assumptions about constant prices be modified to suit the occasion and that the whole approach to the problem of estimating the probable course of consumption, income, employment, and prices be re-examined and revised to take adequate account of all the important factors in the market; (4) that very real limits to what can be accomplished in forecasting by describing the economic world in terms of mathematical relationships

* Paper delivered before the American Statistical Association, Cleveland, Ohio, January 26, 1946.

be recognized and that the importance of selective judgments with respect to particular periods be emphasized; and (5) that, in view of the present uncertain state of the art of forecasting, care should be taken in formulating policy to rely only so far as is necessary on forecasts arrived at by any method.

During the past decade considerable pioneering work has been done in this country in the measurement and forecasting of consumption, and the bibliography of this field is now quite extended—much more so than I supposed when I accepted this assignment. So far, fortunately, crystallization on any single viewpoint has been avoided for the most part, both in measurement and forecasting, and development of new ideas as to objectives and methods has been spirited. About all I hope to do this morning is to remind you of this and to suggest some of the broader issues which seem to merit further attention, particularly issues relating to forecasting.

Measuring consumption. Great impetus was given to measurement in the field of consumption by the marked deterioration in living conditions which occurred in the early 1930's and by the widespread modification in that period of the laissez-faire views of earlier, more prosperous days. Costs of investigation were regarded in many quarters as no impediment—in fact, surveys of all sorts were encouraged as a means of further unbalancing the Federal budget and providing employment. There was keen interest in what could be done by the Government to improve living conditions, and especially the lot of the lowest one-third. The Consumer Purchase Study of 1935–36 provided information on a scale never before attempted in this field, including detailed analysis of the consumption of 60,000 families classified by income group, degree of urbanization, and other characteristics. In some particulars, such as the coverage of high income groups, this study was not wholly satisfactory, but it marked a real turning point in the whole history of the measurement of consumption. Information was obtained not only on dollar expenditures but also on what consumers were able to obtain for their money in real or physical terms.

Since this initial comprehensive survey there have been further studies of a similar sort in 1941 and again in 1944, but the samples have been extremely small and, even though sampling techniques had improved meanwhile, the results have been of only limited value. The war, it may be noted, called for action to restrict consumption rather than to expand it and to conserve labor rather than to provide employment. The gathering of information not directly essential to the war program was generally curtailed.

Meanwhile, since the early 1930's analysis of current economic developments has been extended in new directions. The focus of many recent studies has been development of an integrated system of broad aggregates covering all activities in terms of income flows and expenditures, whereas in the 1920's the effort was to develop measures of changes in strategic points in the economy and not to formalize estimates of changes elsewhere. Development of this new system has led to considerable work in estimating annual consumer expenditures for consumption items and estimates have been made largely on the basis of biennial Census data for production and less frequent Census data on distribution. Work in this field has provided such expenditure estimates over a considerable period of time in quite some detail by commodity groups but not by type of purchaser. The concepts and procedures were determined with a view to fitting the results into the general analysis of gross product, and the matter of physical volume of consumption was not a primary consideration.

For purposes of current analysis it appeared necessary to have data for much shorter periods than years and the Department of Commerce has developed a series on the basis of much more limited data than those available for annual compilations. These figures have certain limitations as a measure of current monthly changes in consumer expenditures and can be interpreted as representing changes in the physical volume of consumption only through a price deflation process. The subject of price deflation, especially for the war period, is still one of sharp debate, and differences in point of view as to proper deflation have contributed to quite different views as to the current level of consumption and the prospect for changes therein.

Changes in living conditions—the physical volume of consumption—received some attention in the planning of the war program and more recently have received important consideration in planning the settlement of war claims and the floating of loans by this country. In this connection a "Special Combined Committee" set up by the Combined Production and Resources Board made a careful study of existing information concerning changes in living conditions during the war period in the United States, the United Kingdom, and Canada and did what they could to make comparisons of absolute levels of consumption in the three countries. The chairman of the group making this study is the chairman of this meeting and any questions on this may well be addressed to him. It may be noted, however, that this study was an attempt to get at changes in living conditions and not simply changes in consumer expenditures; and that it was an attempt to do this by

building up physical volume estimates directly as well as by deflating expenditure series. It was, moreover, an attempt to make international comparisons meaningful through appropriate reclassification of data available in three different countries and through discreet handling of price deflation problems of peculiar difficulty.

Finally, in order to estimate the importance of increased consumer holdings of liquid assets in affecting postwar spending, opinion surveys have been undertaken in an experimental way. It remains to be seen whether intentions to buy can be taken as seriously as intentions to plant.

These observations as to measurement of consumption have been made to indicate some of the many purposes for which information has been sought, some of the many types of information obtained, and some of the broader limitations of available data for purposes of historical analysis and of forecasting. In passing, perhaps apology should be made to certain people for talking as if no one thought of consumption before the Great Depression—particularly to such as Mr. Engel, the 1919 budget study people, the people who initiated the collection of monthly department store sales data in 1919, and the authors of 1,500 family expenditure studies made here and abroad before the Consumer Purchase Study of 1935–36.

Forecasting consumption. At this point the discussion narrows, in one sense, because the forecasting of consumption as a part of the total economic situation is only one of the many purposes for which information about consumption is useful. Nevertheless, the issues involved in forecasting are themselves pretty broad and forecasts at times are important in determining policies, or at least in rationalizing them; the time when forecasts were recognized as of significance only by those who hoped to turn a quick penny in capital markets is long since past. The period since the end of the war has been one in which widespread forecasts of quick deflation may have hastened the process of decontrol and of tax reduction. Forecasters who made such predictions may well ponder whether the influence of their forecasts has been in the right direction and whether there is anything they can do to improve their forecasting methods for the future. They are, as indicated by these sessions, quite aware of the problem. One particular question is how satisfactory estimates of consumption have been and how they might be put on a better basis. In approaching this question it may be recalled that until very recently no attempt had been made to estimate total consumption or to forecast the course of consumption, except in the most general terms; consequently it would be strange indeed if the

results achieved or the methods developed were as good as might be hoped for on the basis of more experience.

One of the most basic facts about the course of consumption in a country developed beyond the stage of recurrent famines is the relatively stable nature of total consumption, in real terms, during most peacetime periods. This is characteristic of total consumption in the sense of consumer purchases and somewhat more so of total consumption in the sense of consumer use of goods, including durable goods they have on hand. (The usual figures, incidentally, relate to a combination of the two; purchases of new automobiles are taken as the measure of automobile consumption while rent payments and certain costs of home ownership, rather than outlays for new houses, are taken as the measure of services rendered by houses.) Such fluctuations as have occurred in consumption—or at least in dollar consumption expenditures—in most past periods have been closely related to changes in income, though considerably less marked.

Observation of such facts about consumption in recent years has led many analysts to the conclusion that in full employment models—a rather recent invention—or in forecasts of actual conditions, figures for consumption could be derived from figures for consumer income after personal taxes, with minor allowances for the influence of factors other than changes in disposable income. Because of the stability of consumption, and the closeness of records for almost every year to one or another indicated relationship between income and consumption, it has been thought that any errors in forecasting consumption would be small, percentagewise. The great size of the consumption component of total expenditures and the great effect of small percentage errors at this point have not received much attention; consumer expenditures in the fourth quarter of 1945 were at an annual rate of 107 billion as compared with 180 for gross product, and ordinarily are a larger part of the total. Nor has adequate attention been given to the impossibility of avoiding the consequences in the consumption estimates of errors in estimates of other components. For "normal" times the analysis of changes in consumption runs in terms of changes in derived consumer demand, with adequate productive capacity assumed.

In compiling forecasts of consumption and production for the immediate postwar period, it has been generally recognized that output of automobiles and other durable goods for consumers would be restricted by technical reconversion problems and that consequently consumer demand could not be regarded to the usual degree as a determining factor in the volume of production in this field. It has also been

recognized that demand for durable consumers goods would be to some extent independent of current income, owing to depletion of consumer stocks and accumulation of consumer buying power.

Production of nondurable goods, however, was expected to be determined largely by demand factors and consumer demand in this field was thought of as being closely related to income. The limitation of production volume through supply factors in a wide range of producer's and consumer's durable goods, moreover, was expected to limit income payments and thereby demand for nondurable consumer goods.

On the basis of these considerations and allowance for a rapid decline in Government expenditures, a reduction of several billion dollars in the annual rate of consumer purchases of nondurable goods from the middle of 1945 to the end of the year was forecast by some last summer. As events turned out, consumer takings of nondurable goods were at an annual, seasonally adjusted rate of 66 billion dollars in the fourth quarter (a figure later revised to 69.5 billion) as compared with 59 billion in the second quarter of the year. The rise occurred at a time when income payments and disposable income were somewhat reduced. Thus, the usual relationship between consumer income and expenditures for nondurable goods did not hold in this period and, moreover, the decline in income was very much less than had been expected.

Now we might brush aside this failure to forecast even the approximate level of consumption outlays with the observation that this was an especially difficult period to forecast by any method. Or, we might interpret the continued strong consumer demand for nondurables as being so temporary that earlier forecasts would be fulfilled—albeit at a somewhat later date than originally suggested. Or, we might take further counsel about the forecasting procedure itself, recognizing that the period immediately ahead now will also be unusual and perhaps of exceptional importance; and that almost no period is of an average sort except some period in the fairly distant future which can not be examined very closely in advance.

My first suggestion is that full recognition should be given—as it has not been heretofore—to the many influences that bear on changes in consumption over time and especially to the varied importance of these different influences in different periods. To regard consumer expenditures, or even consumer expenditures for nondurables, simply as a variable dependent on disposable income is too restrictive, no matter what the formula. It is true that in peacetime many fluctuations in economic activity stem from business decisions about the purchase of

inventories, new plant and equipment, and the like; and that such fluctuations affect consumer incomes and expenditures. It is also true that considerations other than the consumer market may be paramount in such decisions. Almost always, however, these business decisions are made partly with a view to the consumer market. Moreover, changes occur in consumer stocks of automobiles and other durable goods—and also of semidurable goods—which may have important repercussions in the markets for these products, especially if they have not been adequately foreseen. After the end of the war such influences were of great importance and there was also a rapid increase in the civilian population, with the return of several million persons from the armed forces. Returning veterans, moreover, had special demands, chiefly for men's clothing and the like. Consumers generally were in a stronger financial position than would be usual at existing income levels, because of wartime savings which had been only partially devalued by advances in the cost of living. Clearly, in forecasting consumption in such a period, demand for semi-durable as well as durable consumer goods should be regarded as at least semi-autonomous; and demand even for perishables should be estimated partly on the basis of factors other than current income. In 1946 the strength of the consumer's financial position and of his desires and the paucity of his stocks of goods may continue to have a real influence on buying, keeping consumer expenditures for many "nondurables," for example, "out of line" with past relationships and making the problem of price control more serious than many expect it to be. For later, more "normal" years, the point may be less pertinent but, again, the "normal" year, or quarter, may be elusive.

Once the position is accepted that consumer expenditures are autonomous to a significant degree, the estimates of the gross product must be made partly on the basis of independent consumption estimates; and, further, the estimates for capital formation may need to take these consumption estimates into account—unless, perchance, as in the most recent period, they have been based primarily on the estimates of practical capacity rather than of demand, which has been expected to be far in excess of supplies in the field of capital formation. The degree to which consumption expenditures should be regarded not only as autonomous but also as determining certain other items needs to be considered in appraising the particular situation developing at any particular time.

The second suggestion I have to offer is that more attention should be given to the measurement of past and current consumption, especially in real or physical volume terms. Debate about the extent to which

consumer habits may be expected to change under given circumstances could be raised to a considerably higher level if more were known about the actual course of consumption. Last summer, for example, it would have been very pertinent to know how consumption in real terms stood in relation to 1939 and 1941. Had we, as many forecasters thought, not only avoided a "bed-rock" level of consumption but actually increased aggregate civilian consumption by 20 per cent? (And per capita somewhat more?) Or was consumption little, if any, above the 1939 level, as some others thought? Or was it in a range of 7 to 12 per cent (10 to 15 per cent per capita) above 1939 as the Combined Resources Board group concluded about that time with reference to the year 1944? It seems evident that the level was lower in mid-1945 than in 1944. If this is allowed for and if certain debatable downward adjustments are made in the Combined Resources Board figures for 1944, the mid-1945 level for aggregate civilian consumption would appear to be little above the 1939 level and below the 1941 level—with very marked differences, of course, for particular groups. Whatever the correct answer may be as to the level last summer, it seems clear that, if possible, in the future enough work should be done to establish the facts and to avoid such sharp divergencies of view as to the level of consumption prevailing at the beginning of any particular period. And the information at hand should be reasonably accurate by groups of consumption items as well as for the total if the significance of any level for the total is to be properly appraised. This is a matter, it may be added, of considerable importance in estimating the comparative courses of production and consumption in the period immediately ahead and therefore in estimating when the present expansion period will be terminated.

The third suggestion I offer for consideration is that in making forecasts of consumption and the economic situation generally, more thought should be given to changes in prices, costs, and the bargaining process generally. In the predictions framed in gross product analysis terms, note is usually made of the fact that the figures shown for various types of expenditures are in terms of constant 1944 or 1945 prices as the case may be. Note may also be made of the fact that a stable price situation is assumed in making the estimates, at least tentatively until it can be seen how large aggregate demand may be relative to aggregate supply and therefore what the general price situation may be. But right now one of the chief questions which needs to be considered in the preparation of estimates of consumption expenditures and capital formation is what prices will do or will be expected to do. And these questions, in my view, can not be answered by a global comparison of

supply and demand. Prices are made in many markets and price changes are often cumulative. They are made in real estate and other transactions where only transfer payments are included—as well as in markets for currently produced goods to which the gross product table refers. They may be so high as to alter considerably the actual course of activity—they were in 1919–20—and are practically certain to affect the level of income payments, as, for example, when wage rates are increased. At the moment production is interrupted while bargaining over terms of employment goes on. When the various influences of changes in prices and costs and the negotiation thereof are considered, the estimating process for consumption expenditures and other items may be more difficult but should be in some respects more realistic.

It even appears that in some periods the most significant developments may be easier to see when the whole approach is altered and instead of asking how much employment or unemployment there will be, the question asked is what will happen to prices. I hasten to disclaim any desire to play down the importance of employment or activity; rather, I would like to see what additional can be learned from another approach. As one illustration of what I have in mind, the interpretation of recent inventory figures may be mentioned. In the gross product table, inventory changes shown are small, of the order of three billion dollars annually, and their significance is not increased much by allowing for their effect on consumer expenditures. But why is the estimated change no larger than this? It is recognized that producers and distributors are eager to replenish their inventories but it is thought that they can not get the supplies they seek, except over a long period. If that conclusion is right, it means that for a long time pressure to increase prices and to trade in the black market will be great, if price ceilings are held and enforced in the usual trade channels. Within limits, the smaller the increases shown for inventories, the more serious —in a period like this—is the likelihood of further inflationary price advances. And ultimately, I may add, the greater is the danger of unemployment in a period of reaction from a speculative price rise. The same general analysis applies to the net export figures in the gross product table—if exports are small simply because the goods are not available, it means that many markets are continually in a tight position and that price increases can be avoided only by authoritarian measures. If this same analysis is made for residential building, the evidence is clear right now. People are ready to buy many more new houses than can be built in 1946 and prices are continuing to rise—to levels, incidentally, which are far beyond what the cost indexes, with constant

percentage margins for contractors, would suggest. In a period of this sort a figure which shows only how much will be spent for new houses at constant prices or how many homes will be built gives little idea of the demand for new houses and provides no adequate basis for forecasting price developments in that field. Nor does it even contribute properly to an estimate of global demand to be matched against global supply. As I see it, analysis of the importance of price changes, of the way they develop and cumulate, and of the possible methods of forecasting them needs to be developed by methods somewhat different, and in a sense more comprehensive, than those used in the gross product approach.

Clearly, the suggestions offered here imply something about the emphasis to be placed on the development of mathematical devices in forecasting. Without in any way wishing to disparage the use of mathematics at appropriate points, I think it is important to recognize that there are very real limits to what can be accomplished by trying to describe in mathematical terms an economic world in which changes stem from so many influences of such varied significance in different periods. My emphasis—and this is the fourth of the suggestions about forecasting—would be on improving our information (about inventories, investment in plant and the like, as well as about consumption) and developing an analytical approach sufficiently broad and flexible to promote understanding of all the different types of situations which may lie ahead. No formulae or systems of formulae tied tightly to the past (a short past, too, for data of much reliability) can do anywhere near all the job. And efforts to formalize knowledge may divert attention from establishing consistency between estimates of the parts and the facts to establishing internal consistency of a sort within a system of figures. The choice of independent variables, as noted in the discussion of the first point above, may be arbitrary and give rise to misleading results. Perhaps all this is just another way of saying that the judgment of experienced people familiar with developments in economic and political affairs, domestic and international, is one essential element in good forecasting. Earlier I referred to the "art of forecasting"; that phrase was chosen deliberately.

Implied in all this discussion of forecasting consumption and economic activity generally is a conclusion which may be stated as a fifth point: that the art of forecasting is not yet fully developed—or perhaps even that there are narrow limits to what can be done in this direction—and that forecasters should be careful not to claim too much for their present methods. For many purposes, such as the preparation

of budget messages, forecasts must be made currently by the best means available at any particular time and then be defended. But it hardly seems desirable, in discussing the feasibility of a full employment program, to place great emphasis on the improvement made to date in forecasting techniques. Rather, those who support a full employment program, as I do, might better argue that such a program is essential and that, in order to carry out some parts of the program, provision should be made for obtaining much better information than we have and for stimulating efforts to learn more about forecasting.

The future. The final section of this paper was to deal with the future of measurement and forecasting in the field of consumption but the crystal ball is clouded and I see no formulae for predicting such matters. The bibliography on this phase of the problem is fairly short and the opportunities for people of imagination to make real contributions are great.

Even so, the present statistical situation in this field is very much improved over that prevailing immediately after the First World War, at least in absolute terms. What it may be in relation to needs, I leave you to judge. Looking back over the history of the period between the wars, it seems evident that the need for information and analysis in this field was great right along. An understanding of what was happening in consumption as compared with what was happening in capital formation might have provided the basis for better policy in the nineteen twenties, although the temper of the times was such that any advice economists might have given on this subject probably would have been little heeded. In the past decade considerable thought has been given to objectives and methods and there are numerous economists familiar with problems in the measurement and forecasting of consumption. Moreover, there are many more people in the community generally who are aware of such problems. Consequently, there is more chance that the thought of economists may be pertinent and be reflected in action. It is true that recent carefully considered proposals for a large study of family income and a companion study of consumer purchases have not yet been approved in Congress. But this should not be regarded as too discouraging under existing conditions.

The problem of developing satisfactory time series for the physical volume of consumption and for the physical volume of production of goods for consumers is real and will need to be explored carefully. Preliminary investigation of production data, such as has been made at the Reserve Board, and the work of the Combined Resources Board group show that the underlying data available, particularly on a cur-

rent basis, fall considerably short of what might be desired, as, for example, in the field of textiles.

In the international area the occasion for further development of information and analysis in the field of consumption may not appear with the same force as it is likely to in the domestic situation, and with the difficulties of comparison much greater because of much greater cultural differences, the limits on what can be found out are no doubt narrower. Nevertheless, the question of relations between the nations which "have" and those which "have not" is a very basic one, even as the question of distribution of income and consumption within a country is; and sooner or later we may see developments in the field of international statistics on consumption much greater than it now seems reasonable to expect.

Whatever the developments in measurement may be, there will always be forecasting, and forecasting of consumption will be part of the job. There are likely to be further changes in methods as the pressure of events is brought to bear on the work of forecasters and as more consideration is given to the choice of methods for appraising prospects. Sharper distinctions may be drawn in our thinking about various length time-periods and the extent to which forecasts may be useful and feasible for six-month periods, two-year periods, and quarter-centuries. The tendency to overemphasize similarities between particular and average periods and to discount differences may give way gradually. We need not be too much surprised when any particular set of ABC series is superseded. We should continue to avoid crystallization on any narrow view and to develop all approaches appropriate to the wide variety of situations likely to develop.

THE USE OF ADJUSTING FACTORS IN THE ANALYSIS OF DATA WITH DISPROPORTIONATE SUBCLASS NUMBERS*

R. E. Patterson†

This paper presents a new method for the analysis of variance of multiple classifications with unequal subclass numbers. It is believed that the method is simpler and more expeditious than the standard method of fitting constants. The process of adjusting is accomplished by substituting in the following equation:

$$\bar{X}_{ij} - \bar{X}_j + \bar{X} = A_{ij}$$

where $\bar{X}_{ij}$ is the mean of the ith subclass in the jth row or column, $\bar{X}_j$ is the mean of the jth row or column, $\bar{X}$ is the grand mean and A_{ij} is the adjusted mean of the ith subclass in the jth row or column.

The method is based upon the assumption that the weighted sum of squares of the subclass means that are adjusted for the border mean effects is an efficient estimate of the variance due to interaction. Justification of this assumption is indicated by the fact that the difference between the differences of subclass means for a given classification is unchanged by the adjusting process. It is further demonstrated that if a sufficient number of adjustings are carried out the results will be the same as those obtained by the method of fitting constants.

INTRODUCTION

In the field of animal science data are often encountered in which the subclass numbers are neither equal nor proportional. The consequences of this disproportionality in the subclasses have been reviewed by Yates (1933) and Snedecor and Cox (1935). Considerable limitations have been imposed on the interpretations of such data, the solutions being generally considered, at best, only approximations.

One of the oldest methods of analyzing data with unequal numbers is that of correcting or adjusting for differences among the means. By this method two or more groups may be made, on the average, homogeneous for some given effect. The more important second effect can then be studied and compared without the disturbing influence of the corrected or adjusted effect. Although this method has had extensive

* Credit and thanks are due to Professor C. H. Godbey of the Agricultural and Mechanical College of Texas for help in the preparation of this paper for publication.

† Animal husbandman, Division of Range Animal Husbandry, Texas Agricultural Experiment Station, Agricultural and Mechanical College of Texas.

use, especially by workers in the animal sciences, users usually feel obligated to apologize for it even when no other methods are applicable to the data.

The objects of this paper are (1) to show the relationship between the elimination of variance by a method of adjusting and the segregation of variance by the analysis of variance, (2) to demonstrate the use of

TABLE 1

THE NUMBER AND MEAN VALUE FOR MALES AND FEMALES FOR EACH OF FOUR GENERATIONS*

Generations	*Males*		*Females*		*Total*	
	Number	Mean	Number	Mean	Number	Mean
1	15	12	15	6	30	9.00
2	15	10	15	10	30	10.00
3	15	8	15	7	30	7.50
4	15	9	15	5	30	7.00
Total	60	9.75	60	7.00	120	8.375

* Hypothetical data for the purpose of illustration.

TABLE 2

ANALYSIS OF VARIANCE OF THE DATA IN TABLE 1

Source of Variation	*Degrees of Freedom*	*Sum of Squares*	*Mean Square*
Total	119	792.125	
Between generation-sex subclasses	7	568.125	81
Within generation-sex subclasses	112	224.000	2
Between sex means	1	226.875	227
Between generation means	3	170.625	57
Interactions of sex × generation	3	170.625	57

adjusting for segregating the confused variability in a two-way table when the main effects are non-orthogonal caused by disproportionate subclass numbers, and (3) to offer a method of adjusting as a substitute for the more algebraic and lengthy least square method of fitting constants.

ANALYSIS OF VARIANCE AND ADJUSTING IN DATA WITH EQUAL SUBCLASS NUMBERS

In the analysis of a set of data two problems are confronted—one of estimating the variance due to several sources and the other of testing

the significance of these effects producing the variance. In data with equal or proportionate subclass numbers, the method of analysis of variance gives both an estimation of the variance and a test of significance. The method of adjusting herein reported can be used to eliminate the variance for any effect that can be segregated by the analysis of variance. The utility of this property of adjusting shall be made clear when data with the disproportionate subclass numbers are discussed.

For illustrating the elimination of variance by the method of adjusting, the hypothetical data of Table 1 will be used. The results obtained from these data by the usual method of analysis of variance are shown in Table 2.

TABLE 3

SUBCLASS MEANS OF TABLE 1 ADJUSTED FOR DIFFERENCES IN SEX MEANS

Generations	*Males*		*Females*		*Total*	
	Number	Mean	Number	Mean	Number	Mean
1	15	10.625	15	7.375	30	9.00
2	15	8.625	15	11.375	30	10.00
3	15	6.625	15	8.375	30	7.50
4	15	7.625	15	6.375	30	7.00
Total	60	8.375	60	8.375	120	8.375

Adjusted factors for sex: Male = −1.375, Female = +1.375.

TABLE 4

ANALYSIS OF VARIANCE OF THE DATA IN TABLE 3

Source of Variation	*Degrees of Freedom*	*Sum of Squares*	*Mean Square*
Total	119	565.250	
Between generation-sex subclasses	7	341.250	49
Within generation-sex subclasses	112	224.000	2
Between sex means	1	0	0
Between generation means	3	170.625	57
Interactions of sex Xgeneration	3	170.625	57

As noted in Table 1, the two sex means differ from the grand mean by +1.375 for the males and −1.375 for the females. These differences show how much each sex is above or below the average of the entire data. If the individuals within each sex are adjusted or corrected in such way that each sex mean will be equal to the population mean and equal

to each other, there will, of course, no longer be a difference between the sexes. This can be accomplished by the following equation:

$$X_{ij} - \overline{X}_j + \overline{X} = A_{ij} \tag{1}$$

where X_{ij} is the ith individual in the jth row or column, $\overline{X}_j$ is the mean of the jth row or column, $\overline{X}$ is the grand mean, and A_{ij} is the adjusted ith individual in the jth row or column. It then follows that

$$\frac{S(X_{ij} - \overline{X}_j + \overline{X})}{Nj} = \overline{X}. \tag{2}$$

As this is a type of coding that does not affect the variability within the subclasses, it is therefore not necessary to correct each value in the subclass but only their mean:

$$\overline{X}_{ij} - \overline{X}_j + \overline{X} = \overline{A}_{ij} \tag{3}$$

where $\overline{X}_{ij}$ is the mean of the ith subclass in the jth row or column and $\overline{A}_{ij}$ is the corrected mean of the ith subclass in the jth row or column.

TABLE 5

SUBCLASS MEANS OF TABLE 3 ADJUSTED FOR DIFFERENCES BETWEEN GENERATION MEANS

Generations	*Males*		*Females*		*Total*	
	Number	Mean	Number	Mean	Number	Mean
1	15	10.000	15	6.750	30	8.375
2	15	7.000	15	9.750	30	8.375
3	15	7.500	15	9.250	30	8.375
4	15	9.000	15	7.750	30	8.375
Total	60	8.375	60	8.375	120	8.375

Adjusted factors for generations: 1. −0.625; 2. −1.625; 3. +0.875; 4. +1.375.

Such an adjusting procedure was applied to the sex differences in Table 1, giving the resulting values shown in Table 3. When the analysis of variance was applied to these adjusted values, the values of Table 4 were obtained. Comparing Table 2 with Table 4, it is observed that the variance between generations, interaction of generation × sex, and error or within subclasses are the same but the between sexes is zero in Table 4, while the total sum of squares is reduced by an amount equal to between sexes of Table 2. Thus, the variance between sexes has been completely removed without affecting any of the other sources of variation.

Continuing the process of elimination by adjusting, the values of Table 3 were adjusted for differences among generations. The generation adjustments of the subclass means of Table 3 are shown in Table 5. An inspection of this table reveals that there is still no difference between the sex means and now no differences among the generation means. The results of analysis of variance are shown in Table 6. Here, the total sum of squares is reduced by an amount equal to the sum of the between sexes and between generations of Table 2. There remains in the data, therefore, only that variation attributed to the interactions of sex × generation and the within subclass variation.

TABLE 6

ANALYSIS OF VARIANCE OF THE DATA IN TABLE 5

Source of Variation	*Degrees of Freedom*	*Sum of Squares*	*Mean Square*
Total	119	394,625	
Between generation-sex subclasses	7	170,625	24
Within generation-sex subclasses	112	224,000	2
Between sex means	1	0	0
Between generation means	3	0	0
Interactions of sex × generation	3	170,625	57

An interesting feature of interaction is obtained from a comparison of Tables 1 and 5. To illustrate, in Table 1 the difference between the males of generation one and the males of generation two is +2.0, while the difference between the females of generation one and the females of generation two is −4.0. The range between these two differences is 6. This is a type of variation that determines interaction. The differences between the same two pairs of subclass means in Table 5 are +3.0 and −3.0. The range between the two differences is 6.0 as before. This shows that interaction has not been affected by adjusting for the main effects. Although the differences may be changed, the range between the differences will always remain the same. This will be considered again in connection with disproportionate subclass numbers.

The interactions of sex × generation also can be removed from Table 5, leaving the subclass means equal and equal to the grand mean. The removal of interaction can be accomplished before or after either of the main effects are removed by adjusting for each of three correctly chosen sets of means.

If the analysis of variance is made considering the main effects eliminated, the corresponding degrees of freedom for those effects must be

considered utilized in the adjusting and removed from the total. Thus, in the analysis of Table 4 where the sex effect was eliminated there would be 118 degrees for total, three degrees for between generations, three degrees for interaction, and 112 degrees for within subclasses, the one degree of freedom for between sexes being utilized in the adjusting. Ignoring the reduction in degree of freedom when there has been a reduction in sum of squares by adjusting might result in an erroneous conclusion from the data.

ADJUSTING IN DATA WITH DISPROPORTIONAL SUBCLASS NUMBERS

In data where the subclass numbers are unequal and disproportionate the ordinary method of analysis of variance is not applicable. Snedecor and Cox (1935) have pointed out some of the difficulties encountered in such data. The principal one is that the addition theorem does not

TABLE 7

DATA TAKEN FROM SNEDECOR AND COX (1033)

Generations	*Males*		*Females*		*Total**	
	Number	Mean	Number	Mean	Number	Mean
1	21	76.952	27	0.518	48	30.020375
2	15	61.467	25	14.080	40	31.850125
3	12	55.667	23	8.522	35	24.680000
4	7	71.000	10	6.700	26	24.077308
Total*	55	07.327291	04	0.030101	149	31.120826

* The means for the border totals are carried six places in order to insure a more accurate comparison with the values obtained by the least square method of fitting constants.

apply. It is desirable to point out other direct consequences of disproportionate subclass numbers upon the means of the main effects as well as upon the interactions between these main effects. For discussion and re-analysis the data given by Snedecor and Cox (1935) will be used. These are shown in Table 7, where each of the means is decreased by 100.

There is no question that the subclass means of a set of data are good estimates of the parameter subclass means, even when the numbers are disproportional. However, when the subclass numbers are disproportional, the differences among the border means are not true estimates of the parameter differences, because these differences are determined not only by the effects of that classification but include also some of the effects that are exhibited in the other or other classifications. In other

words, the main effects are not orthogonal, due to disproportional subclass numbers. In the example the difference between the sex means is no single function of sex but also reflects some of the generation differences. Likewise, the differences among the generation means not only reflect all the effects of generations but a part of these differences is caused by differences between sexes. Fundamentally, this is the primary reason the addition theorem does not hold—the main effects are confounded or non-orthogonal.

Of the several methods that have been developed for the analysis of data with disproportionate subclass numbers, the least square method of fitting constants, as reported by Brandt (1933) and extended in its

TABLE 8

DATA OF TABLE 7 ADJUSTED TO ZERO SEX DIFFERENCE*

Generations	*Males*		*Females*		*Total*	
	Number	Mean	Number	Mean	Number	Mean
1	21	40.745535	27	30.702035	48	35.000404
2	15	25.260535	25	36.261035	40	31.613098
3	12	19.460535	23	20.706035	35	20.193086
4	7	34.703535	10	27.074035	20	29.810403
Total	55	31.120820	94	31.120820	149	31.120820

Adjusting factors for sex differences: Male = −30.206465, Female = +21.184035.

* Example, to compute 40.745535, the mean of the males in generation one, add −30.120820 to 70.052 of Table 7.

application by Yates (1934), is considered to have the widest range of application and the most accurate results. The primary assumption is that there are no interactions in the population from which the sample is drawn. In this method a set of constants is fitted to the data with the condition that they determine a set of subclass means with zero interactions but with the values of the main effects unchanged. Next is calculated the reduction in sum of squares due to fitting the constants, and from this an appropriate correction for disproportionate subclass numbers is obtained.

By the method of adjusting, practically the same results are obtained as by the method of fitting constants. With this method the assumption is made that the sum of squares of the subclass means that are adjusted for the border effects is an efficient estimate of the variance due to interaction. If the difference between differences of subclass means for a given classification determines interaction, the assumption is highly

justified. That is, the border effects are not confused with the interaction between these effects.

By using the data in Table 7, the results of the method of adjusting are compared with those obtained by Snedecor and Cox (1935). Consider that the first object is to estimate the sum of squares due to generation means. This is to be based on generation means stripped of any influence of sex differences. Therefore we shall remove that variation due to sex by substituting in equation 3. This gives a set of subclass means that are not influenced by sex (Table 8). However, as the sex means had some influence of generations, as described above, the adjusting removed this also. When the analysis of variance is applied, the weighted sum of squares between generations is found to be 1659.13869. This is smaller than it should be for the reason just given.

TABLE 9

DATA OF TABLE 8 ADJUSTED TO ZERO GENERATION DIFFERENCES

Generation	*Male*		*Female*		*Total*	
	Number	Mean	Number	Mean	Number	Mean
1	21	30.760057	27	28.727057	48	31.120826
2	15	24.808203	25	34.872303	40	31.120820
3	12	24.387075	23	34.633775	35	31.120820
4	7	30.103868	10	29.284008	26	31.120820
Total	55	30.737677	94	31.345000	149	31.120820

Adjusting factors for generation differences of Table 8: 1. −3.075576, 2. −0.302272, 3. +4.027140, 4. +1.310333.

In order to recover some of the generation effects that were removed with sex because of the disproportionate subclass numbers, the subclass means in Table 8 are adjusted for generation effects. Because of the disproportionate subclass numbers, the difference between the sex means will no longer be zero (Table 9). By calculating the between sex weighted sum of squares, 12.80149 is obtained. This value is part of the generation sum of squares removed with sex in the initial adjusting. It is perhaps well to emphasize here that the final resulting sum of squares will be the result of contributions from both sets of border means. Adding 12.80149 to 1659.13869 gives the sum of squares for between generations of 1671.94018, a value within the limits of rounding equal to that found by the method of fitting constants.

The next step is to readjust for sex effects. When this is done, the weighted sum of squares between generations is found to be 0.19049.

As this process is continued, each additional correction will recover a part of the between generation sum of squares that was eliminated with sex initially. In this example, however, the amounts recovered after the first two adjustments are very small. The sum of squares for between generation, as found in six adjustments, is 1672.14380 or the sum 1659.13869 + 12.80149 + 0.19049 + 0.00712 + 0.00371 + 0.00230, representing the first to the sixth adjusting respectively.

TABLE 10

DATA OF TABLE 7 AFTER SIX ADJUSTMENTS, THE INITIAL ADJUSTMENT BEING FOR SEX DIFFERENCES

Generation	*Male*		*Female*		*Total*	
	Number	Mean	Number	Mean	Number	Mean
1	21	37.110607	27	26.457440	48	31.120826
2	15	25.253418	25	34.641270	40	31.120826
3	12	24.702037	23	34.422480	35	31.120826
4	7	36.554203	19	29.119055	26	31.120826
Total	55	31.120743	94	31.120874	149	31.120826

TABLE 11

DATA OF TABLE 7 ADJUSTED TO ZERO GENERATION DIFFERENCES

Generation	*Male*		*Female*		*Total*	
	Number	Mean	Number	Mean	Number	Mean
1	21	60.052541	27	1.618451	48	31.120826
2	15	60.737701	25	13.350701	40	31.120826
3	12	62.101826	23	14.056826	35	31.120826
4	7	78.043518	19	13.833518	26	31.120826
Total	55	60.412610	94	10.471378	149	31.120826

Adjusting factors for generation differences of Table 7: 1. −7.800540, 2. −0.720200, 3. +6.434826, 4. +7.043518.

Table 10 shows the data after the process of elimination is continued six times, beginning first with the elimination of the sex differences. As may be observed in this table, each border mean is approximately equal to the grand mean. The variability among the subclasses is therefore considered due to the effects of interactions. By calculating the weighted sum of squares for between subclasses, a value of 3182.40355 is obtained. This is, within the limits of rounding, equal to the interactions found by the least square method of fitting constants.

By the same procedure the sum of squares for between sex means can be obtained. After adjusting for generation effects of Table 7, the sum of squares for sex (Table 11) was found to be 108,584.62490. When adjusted for sex differences the sum of squares for between generations (Table 12) was 1594.09223, which is actually due to sex. Then 108,584.62490 plus 1594.09223 gives 110,178.71713. This value is still somewhat smaller than that found by the method of fitting constants.

TABLE 12

DATA OF TABLE 11 ADJUSTED TO ZERO SEX DIFFERENCES

Generation	*Male*		*Female*		*Total*	
	Number	Mean	Number	Mean	Number	Mean
1	21	33.760757	27	22.267809	48	27.265985
2	15	25.445917	25	34.000140	40	30.702312
3	12	26.810042	23	35.600274	35	32.590423
4	7	42.751734	19	34.482060	26	30.700179
Total	55	31.120826	94	31.120826	149	31.120826

Adjusting factors for sex differences of Table 11: Males = −35.291784, Females = +20.649448.

TABLE 13

DATA OF TABLE 12 ADJUSTED TO ZERO GENERATION DIFFERENCES

Generation	*Male*		*Female*		*Total*	
	Number	Mean	Number	Mean	Number	Mean
1	21	37.585508	27	26.092740	48	31.120826
2	15	25.774431	25	34.328603	40	31.120826
3	12	25.340445	23	34.130677	35	31.120826
4	7	37.163387	19	28.894619	26	31.120826
Total	55	31.638065	94	30.817680	194	31.120826

Adjusting factors for generation differences of Table 12: 1. +3.824841, 2. +0.328514, 3. −1.469507, 4. −5.586347.

However, a considerable amount of sex influence was removed with generations and not all was recovered at the first readjustment. Readjusting for generation differences (Table 13) yields a sum of squares between sexes equal to 23.52606. This is also due to between sexes and when added to 110,178.71713, gives 110,202.24319, a value similar to that found by fitting constants. The next adjusting would yield a sum of squares less than 0.50. There was thus needed two readjustments to regain most of the sum of squares for sex that was eliminated with

generation in the initial adjustment. If the elimination process is continued, a set of subclass values would be obtained that are practically equal to those of Table 10, where the initial adjustment removed the sex effect.

In actual practice it is necessary only to find the unaffected adjusted sum of squares for one of the border effects, as the difference between the unadjusted and adjusted sum of squares for one effect is equal to the same type of difference for the other border effect. Thus, the difference between the original sum of squares for generation (5756) and the adjusted (1672) equals 4084. To find the sum of squares for between sexes subtract 4084 from the unadjusted sum of squares (114,287), which gives 110,203, as obtained above for between sexes. Interaction may also be calculated from the subclasses that have been adjusted for either set of border means in the initial adjustment.

Table 14 shows the combined results of the analysis of variance by the method of adjusting. The mean square values when rounded off

TABLE 14

ANALYSIS OF VARIANCE OF TABLE 7 BY THE METHOD OF ADJUSTING

Source of Variation	*Degrees of Freedom*	*Sum of Squares*	*Mean Square*
Between sex means	1	110,202.12	110,202
Between generation means	3	1,672.03	557
Interactions	3	3,182.89	1,061
Within subclasses	141		400

are equal to those found by Snedecor and Cox (1935) from fitting constants to the same set of data. The sums of squares by the two methods differ because of rounding off and because a part of the sum of squares removed in each initial adjusting was never recovered.

It will be observed that the differences between the differences of the subclass means of Table 10 are the same for corresponding values found in Table 7. For example, the difference between the two male subclasses for generations one and two in the unadjusted data is +15.485, while that between the female subclasses for generations one and two is −5.562. The difference between these two differences is 20.047. In Table 10 the difference between the corresponding first pair of subclasses is 11.863179 and between the second pair is −8.183821. The difference between these two differences is 20.047000, the same as above for the unadjusted data. If interaction is determined by such

differences, adjusting for the border means has in no way affected the interactions.

Furthermore, the method of adjusting applied to data computed from fitted constants, so that the interactions are zero, gives the same estimates of sums of squares for the main effects as those obtained when the original data are used. Thus, the sizes of the estimates of sums of squares for the main effects are not influenced by the amount of interaction present. There is also no evidence that the variances for the border effects are less well estimated as the departure of the subclass numbers from proportionality becomes greater. In such cases, however, the border effects become more confused or confounded.

The adjusted border means may also be calculated by the method of adjusting. Such calculated means are devoid of the confounding effect brought about by disproportionate subclass numbers. The adjusted mean of generation one is found, for example, by adding to the sex adjusted generation mean the difference between the succeeding sex adjusted means and the grand mean. Thus, the original sex adjusted generation one mean is 35.09640. The same generation after adjusting for sex the second time is 31.16235, which differs from the grand mean by 0.04152. The third sex adjusted generation one mean is 31.12144 with a difference of 0.00061. Further sex adjustments will result in continually small differences. The mean for generation one will be for the three sex adjustings 35.09640+0.04152+0.00061 =35.13853. In a similar manner, generation two is found as 31.51672. The difference between these two means is 3.62181. This is in very close agreement with the difference between the two (b) constants for the same two generations. By the same method, adjusted sex means can be obtained as well as the difference between sexes unaffected by generation differences.

Both sets of adjusted border means can be found when the initial adjustment has been for either one of the effects. In other words, the sex means can be calculated when the initial correction was for the sex difference. For example, the sum of the unadjusted mean for the males (67.327291) and the differences between sex and generation adjusted male means and the grand mean (−0.38149 plus −0.005625) equals 66.938517. By the same method the mean for the females was found to be 10.163665. The difference between the two sexes is therefore, 56.774852, a difference similar to that between the (a) constants of the least square method of fitting constants. Thus, it is necessary only to carry out the elimination process for one set of border means and then make as many recovery adjustments as may be considered necessary.

SUMMARY AND CONCLUSIONS

A method of adjusting is described whereby the sum of squares of the various sources of variance can be eliminated. When this method of adjusting is applied to data with unequal subclass numbers, it is possible to obtain a sum of squares for each source of variance that is free of the influence of the other effect.

The method of adjusting is applied to data with unequal subclasses that had been analyzed by the least square method of fitting constants, showing that the same results can be obtained by both methods. It is indicated that the adjusting method is much simpler and requires less laborious mathematical computations than the method of fitting constants and is therefore offered as a substitute for it.

Although the method of adjusting has not been tested extensively nor subjected to algebraic proof, it has given results similar to those obtained by the method of fitting constants in several two-way sets of data.[1] It seems safe, therefore, to conclude that the method can be substituted for the least square method in data where the latter is appropriate.

REFERENCES

Brandt, A. E. (1933). The analysis of variance in a $2\times s$ table with disproportionate frequencies. *Jour. Am. Statistical Assoc.*, Vol. 28, p. 164.

Snedecor, G. W. and Cox, G. M. (1935). Disproportionate subclass numbers in tables of multiple classification. *Research Bul. No. 180*, p. 235. Ames, Iowa.

Yates, F. (1933). The principle of orthogonality and confounding in replicated experiments. *Journal of Agri. Science*, Vol. 23, p. 108.

Yates, F (1934). The analysis of multiple classifications with unequal numbers in the different classes. *Journal Am. Statistical Assoc.*, Vol. 29, p. 51.

[1] W. G. Cochran informs me that it can be proved algebraically that if a sufficient number of adjustments are carried out, the method of adjusting gives exactly the same results as the method of fitting constants.

A PUNCHED CARD METHOD FOR PRESENTING, ANALYZING, AND COMPARING MANY SERIES OF STATISTICS FOR AREAS

BERTRAM J. BLACK
Former Director, Bureau of Social Research
Federation of Social Agencies of Pittsburgh and Allegheny County

AND

EDWARD B. OLDS
Director, Research Bureau
Social Planning Council of St. Louis and St. Louis County

Information about small areas such as census tracts, wards, cities, counties, etc., should be available to the consumer in a compact readable form. To meet the problem of multitudinous tables in large expensive volumes, the method described below was devised for the analysis and publication of data concerning the neighborhoods of Pittsburgh and Allegheny County. By using punched card procedures, making use of summary cards, machine multiplication and other techniques, a set of cards was prepared containing the social characteristics of each area and including percentile rankings identifying the relative position of each area for a particular characteristic. The statistical tables were reproduced directly from the listing sheets. Instead of repeating identical heading stubs for each of the tables, one set of stubs was printed on overlapping tab cards bound to the right side of the book. This alone makes for a most efficient use of paper and reduces the size and cost of the books. This method of analyzing and presentation can be applied to data for cities, counties, and metropolitan areas, as well as to small areas within cities.

THE increasing quantity of available information for small areas such as census tracts, wards, study areas, cities, and counties is making the work of analysis, comparison, and publication more and more complicated and time-consuming. The consumer of statistics for a particular area does not usually have the time nor the inclination to spend weeks calculating rates and percentages so as to determine the standing of his city or neighborhood among others. Neither does he wish to consult a dozen different sources for his basic information. The data should be available to him in a compact, readable form so analyzed that he can see how his community compares with others in the most important characteristics for which statistics have been compiled.

This problem has only partially been solved by the publication of tables which show percentages, rates, and rankings for a group of cities, counties or census tracts. Physical limitations upon the size of tables

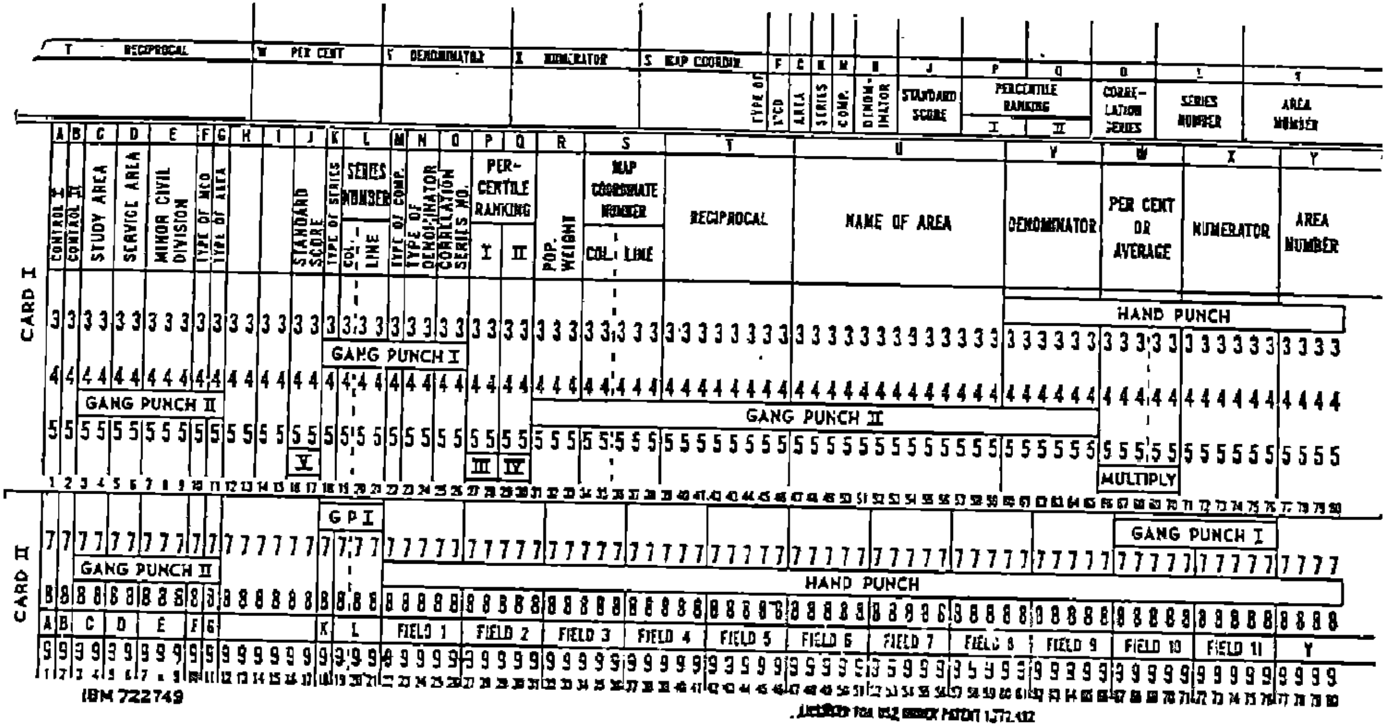


FIG. 1. THIS CARD IS DESIGNED TO SERVE AS TWO CARDS. THE HEADINGS ON THE TOP OF THE CARD REFER TO CARD I; THE CAPTIONS TOWARDS THE BOTTOM OF THE CARD REFER TO CARD II. CARD I IS DESIGNED FOR THE MAJOR PART OF THE PROCEDURES; CARD II IS DESIGNED PRIMARILY FOR THE TABULATION OF FREQUENCY DISTRIBUTIONS.

necessarily prohibit the presentation of more than about half a dozen series in one table. To obtain information on several hundred series, it is necessary to search scores of tables. This necessarily requires a large, cumbersome expensive volume.

Another approach to the problem has been attempted through the publication of specially prepared booklets for each area. This is an ideal solution if funds are available to cover the cost of publication. Unless the areas are large counties or cities, it is doubtful whether the returns from sales can be expected to cover costs.

The method of meeting the problem outlined in this paper has been applied to the analysis and publication of data concerning the neighborhoods of Pittsburgh and Allegheny County. The techniques used grew out of experience gained over the preceding decade in the analysis of small area data and in their presentation. Through the assistance of NYA and WPA projects during the Depression years, many special tabulations had been made of local data.[1] Most of these data had been summarized in a series of "Social Facts Booklets." These consisted of mimeographed forms on which the data were typed. The preparation of copies of these booklets proved to be slow and expensive. Many users were interested in more than one booklet which further complicated the problem of distribution.

The vast quantity of new data from the 1940 Population and Housing Census which needed to be analyzed and compared with the 1930 census data, as well as with data from numerous local tabulations, presented a real problem in clerical man hours. By the time the data were available, clerks and calculating machines were difficult to find. The choice seemed to narrow down to postponing extensive analyses until after the war or of developing some labor saving method of doing the job. Because of the great interest in the data for post-war planning purposes, it did not seem wise to postpone. Accordingly, considerable work was done in the summer of 1942 in the development of a plan for the analysis and publication of the material through the use of punched cards.

Since the basic data were available in most instances by census tracts, the plan provided for the key punching of one card for each census tract for each series to be analyzed. The basic card form used is illustrated in Fig. 1 (Card I, the upper half of the card headings). The design of this card was changed many times during the course of laying out the plan for the project. Only the last 21 columns of this card were

[1] Edward B. Olds, "The Use of NYA Workers in Ecological Studies," *Social Forces*, Vol. 20, No. 2, December, 1941, pp. 218–223.

manually punched and in some instances only the last 10 columns. These fields showing the denominator or base used in computing percentages or rates, the actual per cent, rate or average, the numerator or frequency for the particular area for the particular series, and the area code number, were punched directly from the published census tract figures wherever possible, or from carefully prepared work sheets. The fields labelled "Gang Punch I" were then punched by the reproducing machine to indicate the various codes for the particular series. These codes will be described later. After all the cards were punched and verified for the various series, the cards were sorted by area number. Previously punched master cards were then used to reproduce area codes and descriptions into the fields labelled "Gang Punch II."

Another type of card form shown in Fig. 1 (Card II, the lower half of the card headings) was used to obtain summaries of frequency distributions for such series as "Monthly Rent of Homes." To compute medians for areas made up of two or more census tracts, it was necessary to obtain totals for each class of the frequencies for each tract. Columns 3 to 11 on this card form contained the same area code information as the Card I form on the upper half of Fig. 1.

The area codes were used to indicate the code number of the embracing study area, service area, or minor civil division for the particular census tract. By sorting on these codes, it was possible mechanically to punch summary cards for the various types of summary areas. In this manner it was not necessary to manually add or to key punch the cards for summary areas. Since the first two digits of the Pittsburgh census tracts indicate the Ward number, it was possible to obtain summary cards for wards by sorting and controlling on this number. The field labelled "type of MCD" was used to designate whether the minor civil division was a Borough, Township, or City and whether or not the area fell within the census definition of an urban place. The field labelled "Type of Area" indicated whether the area was a census tract, ward, study area, service area, minor civil division, county service area, or county ward. These various types of areas had developed historically as the most useful areas for purposes of neighborhood and community planning.

The population weight for each area (Field R) was arrived at by dividing the 1940 population of the area by 100. It was used as the weight of the area in computing the percentile ranking (to be described later).

The Map Coordinate Number was used to indicate the exact position of each area on the census tract base map for Allegheny County.

This map was drawn with the possibility of running it through a tabulator and to have decile rankings listed directly on the map on each census tract as the basis for subsequent cross hatching. However, this method was found to be impractical in actual practice, and the maps were manually prepared from straight machine listings.

The reciprocal of the denominator was gang punched in along with the area codes where the same base was applicable to a number of series. For example, the total population of an area was used over and over to compute percentages and rates. By punching the reciprocal of population into the cards for all these series, it was possible to compute the percentages by machine multiplication. For this a multiplier is used.

The name of the area was gang punched into each card to provide easy identification when the cards for each series were later listed in order by the percentages. This listing showed the names of the areas that were highest and lowest as well as all the shades in between.

It was decided to use "percentile rankings" to identify the relative position of each area in the array for a particular series. This seemed superior to simple ranks because it was independent of the number of areas in the array. For the publications herein described, the number of areas varied as follows:

Minor civil divisions	126
County consers tracts	297
County service areas	26
Pittsburgh service areas	19
Pittsburgh study areas	53
Pittsburgh wards	32
Pittsburgh census tracts	194

In addition, there were different numbers of minor civil divisions and of census tracts in 1930 and 1940. By reducing all ranks to percentile ranks, it was possible to indicate the standing of an area in a particular series without reference to these differences.

The percentile ranks can be used to divide the areas into other desired groupings. For example, to divide the Pittsburgh census tracts into thirds on the basis of some factor, such as average rents, any tract with a percentile rank between 0 and 33 is in the lowest third; between 34 and 67 in the middle third; and between 68 and 100 in the upper third. Since the percentile rank has been weighted to allow for the population of each tract, it can be used to divide the population into thirds, fourths, fifths, sixths, eighths, tenths, etc., on the basis of rates or percentages for neighborhoods.

	[illegible]	TYPE OF AREA	NAME OF AREA	COUNTY CENSUS TRACT NUMBER	NUMERATOR	DENOMINATOR	PER CENT OF AVERAGE	STANDARD SCORE	PERCENTILE I	PERCENTILE II	CORRELATION SERIES NUMBER
2703	301	0	PGH 27C				332		76	72	50
4130	301	0	MCCANDLESS T	130			332		77		50
4153	301	0	MCKSPORT CITY	153			332		77		50
4253	301	0	STOWE T	253			332		77		50
1304	301	0	PGH 13D				331		77	72	50
2206	301	0	PGH 22F				331		77	73	50
2804	301	0	PGH 28D				331		77	73	50
4082	301	0	EDGEWOOD B	082			331		78		50
4151	301	0	MCKSPORT CITY	151			331		78		50
4284	301	0	WEST VIEW B	284			331		78		50
4289	301	0	WILKINSBURG B	289			331		78		50
1907	301	0	PGH 19G				330		79	74	50
4133	301	0	MCKEE ROCKS B	133			330		79		50
4173	301	0	MILLVALE B	173			330		79		50
305	301	0	PGH 3E				329		79	74	50
4181	301	0	MT OLIVER B	181			329		79		50
2806	301	0	PGH 28F				328		79	74	50
4180	301	0	MT OLIVER B	180			328		80		50
4236	301	0	SEWICKLEY B	236			328		80		50
4294	301	0	WILKINSBURG B	294			328		80		50
1108	301	0	PGH 11H				327		80	75	50

FIG. 2

In computing the percentile ranks, the population weights for the areas in the array were progressively totaled and printed on the tabulated list. This list was then marked off into 100 parts each of which included approximately one per cent of the total population. The punched cards were gang punched to indicate the appropriate percentile rank on each card. The cards were then sorted into order by area and listed so as to indicate the significant information, including the percentile rank for each area, as illustrated in Fig. 2.

After percentages, rates, averages and percentile ranks were computed for all areas, the cards were assembled and sorted to prepare the summary reports for each area for all series. These area reports were considered the final reports from the punched cards and were used as the copy for the photo-offsetted sheets bound in four volumes. Instead of repeating identical heading stubs for each of the numerous area summary tables, one set of stubs is printed on overlapping tab cards bound to the right side of the book. Each of the six overlapping tab cards matches with one of six columns of data on the area tables which are bound to the left side of the book. This arrangement makes possible the most efficient possible use of paper and cuts down the size and cost of the books. It also facilitates the rapid location of any desired information for any area in the county. The tabs are illustrated in Fig. 3.

Three types of information are presented in each column of the table. These consist of the actual frequency for the particular series such as the number of foreign-born persons; the percentage, such as the per cent of the total population which was foreign-born; and the percentile rank of the area according to the per cent foreign born. This detail is shown for approximately 175 series of data in the city of Pittsburgh. The bottom part of each table was used to show frequency distributions without any percentages or percentile ranks for such series as population by age and sex, persons per household, monthly rent of homes, years of school completed, major occupation group, and type of structure. Selected information from these series was translated into rates, percentages, or averages and was presented also in the upper part of each table.

The preparation of the area reports was greatly facilitated by the use of the punched cards to run lists which were photo-offsetted directly, thereby minimizing typographical errors. The series code gang punched into the cards immediately after key punching, was so set up as to place the cards correctly on the final area tables. Thus series 101 was listed on the first line of the first column on the report; series 102

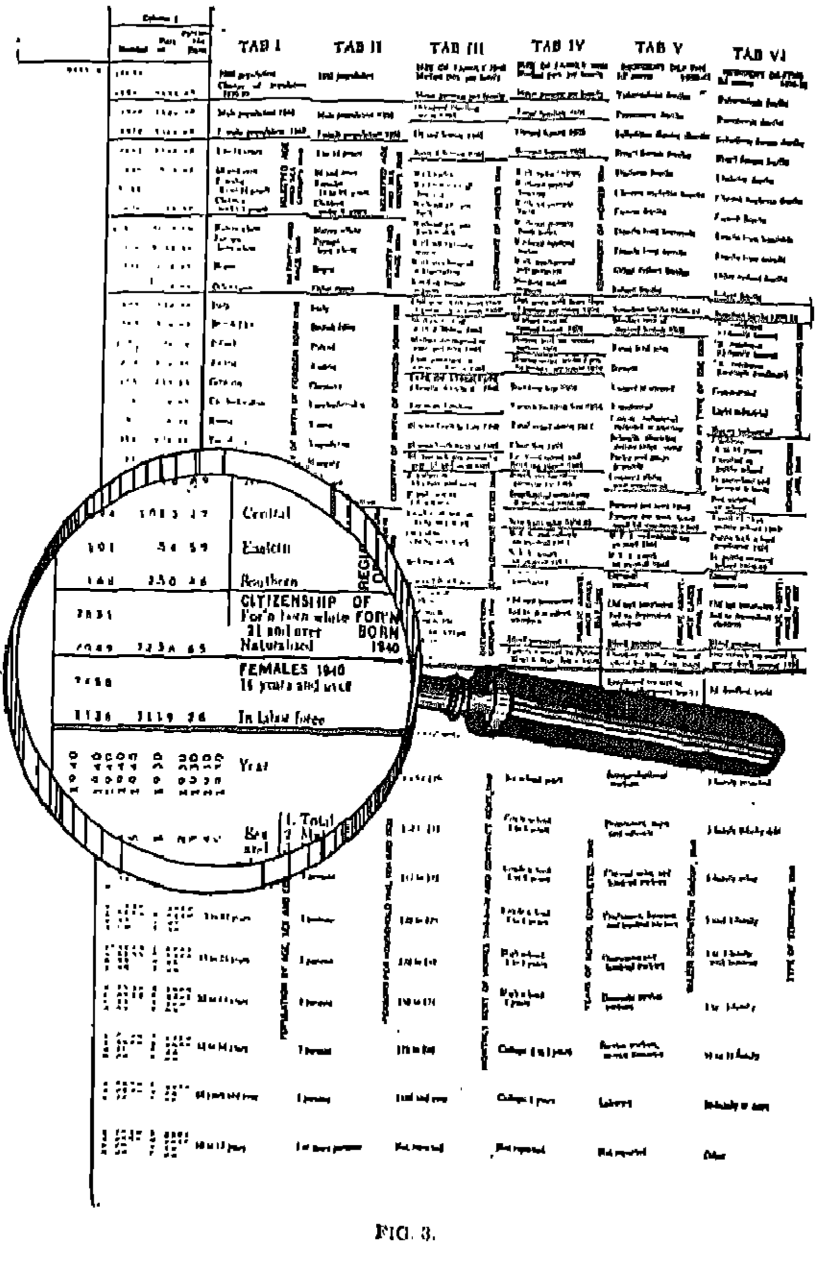

TAB I
TAB II
TAB III
TAB IV
TAB V
TAB VI
Central
Eastern
Southern
CITIZENSHIP OF
For'n born white FOR'N
21 and over BORN
Naturalized 1940
FEMALES 1940
14 years and over
In labor force
Year

FIG. 3.

on the second line of the first column; series 203 on the third line of the second column, etc. The final copy for the printer was obtained by pasting together two listings for the upper part of the table together with the listing for the lower part. Decimal points had to be put in by hand and some touch-up work was required. Guide lines were printed in red on the final tables to facilitate locating individual series without reference to the tabs.

Considerable interest has been shown in the final volumes by planning agencies, real estate companies, schools, newspapers, market analysts, and social agencies. One volume contains the ward information for Pittsburgh; another volume contains the study area information for Pittsburgh; the other two volumes are devoted to the 126 minor civil divisions of Allegheny County. It is expected that the cost of publication will be covered by the sale of the books. The cost of the analytical work was covered by a special grant from the Buhl Foundation.

The basic cards are being retained for special tabulations and listings for various studies which may be desired in the future. For example, it may be of significance to make summary tabulations of several series of data for somewhat different combinations of census tracts than those which have been used in the past. To obtain such tabulations, it is a relatively simple machine procedure to reproduce the census tract cards for the desired series and to gang punch new area codes. The cards can then be sorted by the new area codes and tabulated to obtain the totals for the new combinations of census tracts. The cards will also be available for correlation studies to demonstrate the close relationship between such factors as average rents and the birth rate, poor housing and large families.

It is believed that this method of analysis and presentation could be applied to data for cities in various size classes or regions, to counties, and to metropolitan areas, as well as to small areas within cities. The chief difficulties encountered resulted from the constant turnover in the personnel engaged in the undertaking, as well as from problems of obtaining sufficient tabulating machine facilities. Both of these difficulties were unavoidable during the war period but would not be appreciable factors in times of peace.

LINEAR REGRESSION FUNCTIONS WITH NEGLECTED VARIABLES*

HOWARD L. JONES
Illinois Bell Telephone Company

This article discusses some properties of the computed Y values obtained by fitting a linear regression function to independent observations by the method of least squares. For the general case where the form of the fitted function may not be correct it is proved that (a) the sampling variance of the computed values and of the residual differences is the same as for the special case where the form of the fitted function is correct, and (b) the mean square bias of the set of computed values is less than, or equal to, that of any other set of linear estimates. These and other properties lead to the suggestion that in minimizing the mean square error, one or more variables be neglected unless Snedecor's F is greater than two.

IN MAKING forecasts on the basis of empirical regression functions fitted by the method of least squares, it is a common experience to find the forecasting errors to be so large that they can hardly be ascribed to chance alone. In fitting such functions, the following possibilities should therefore be considered:

1. The assumption about the errors of observation that underlies the least square procedure may not be realistic. In particular, the errors in the dependent variable may be serially correlated.
2. The assumption about the relationship between the variables may not be correct. In particular, the complete relationship may involve variables not included in the regression function, or the form of the relationship between the variables may differ from the form of the fitted function.

There is no satisfactory method now available for determining, from internal evidence, that one of these two difficulties, but not the other, is present. In fact, the two difficulties may both be present. By referring to past experience, however, some indication may be had as to which one is likely to be the principal source of trouble.

Consider, for example, the problem of fitting n terms in a time series by using a k-parameter polynomial function of time,

* This article is the result of a search for a criterion to use in making a choice among several linear estimates of trend ordinates in a time series. Among these estimates were some computed from biased functions, developed jointly by John H. Smith and myself, in which least square regression functions were treated as independent variables in trying to find the linear function that minimizes the expected value of the mean square error. It is a pleasure to acknowledge my debt to Dr. Smith for his encouragement at all times.

$$f = a_0 + a_1 t + a_2 t^2 + \cdots + a_{k-1} t^{k-1}.$$

Past experience may indicate that a satisfactory fit to a given set of observations can usually be obtained; but that when the polynomial is projected much beyond the range of the original observations the fit becomes unsatisfactory. In this situation, suppose it usually happens that the poorness of fit is due solely to the computed values of the parameters and that, no matter how many homogeneous terms are added to the series, a satisfactory fit to the augmented series can still be obtained with a polynomial of the same degree $(k-1)$ by making suitable adjustments of the original parameters. Then the principal source of difficulty may reasonably be ascribed to the assumption about the errors of observation. On the other hand, if it is usually necessary to increase the degree of the polynomial in order to obtain a satisfactory fit to a longer series of the same kind of data, then at least part of the difficulty in fitting these regression functions may be ascribable to the assumption about the relationship between the variables.

In this article, the mean square error and some other properties of linear regression functions will be investigated for the case where the second difficulty, but not the first, may be present.[1] For this purpose, the mean square error is defined to be the mean of the squared differences, at the points actually observed, between the estimates computed from the regression function and the expected values of the dependent variable. For the general case where the form of the fitted function may not be correct, the following results will be proved:

1. The mean square error of the estimates is the sum of two components, the mean sampling variance and a component which may be called the mean square bias.
2. The sampling variance of the estimates, and of their differences from the observed values of the dependent variable, is the same for the general case as for the special case where the form of the fitted regression function is correct.
3. The mean square bias of the estimates is less than, or equal to, the mean square bias of any other linear function of the same independent variables.

[1] For a discussion of the mean square error when neither difficulty is present, see Markoff, A. A., *Wahrscheinlichkeitsrechnung*, 1912, pp. 201–228 (translated into German by H. Liebmann from the Russian 2nd ed.); also David, F. N., and Neyman, J., "Extension of the Markoff Theorem on Least Squares," *Statistical Research Memoirs*, vol. 2 (1938), p. 105. For the case where the errors in the dependent variable are serially correlated, see Aitken, A. C., "On Fitting Polynomials to Data with Weighted and Correlated Errors," *Proceedings of the Royal Society of Edinburgh*, vol. 54 (1933), p. 12; also, Dixon, W. J., "Further Contributions to the Problem of Serial Correlation," *Annals of Mathematical Statistics*, vol. 15 (1944), p. 110, and references cited on p. 144.

In the section headed "Practical Applications," suggestions will be offered concerning the use of these properties in actual problems, including the problem of making a choice among several linear regression functions where minimum mean square error is a satisfactory criterion. Employing the suggested criterion turns out to be equivalent to applying a test of significance to the partial regression coefficients and then neglecting the corresponding independent variables unless Snedecor's F is greater than two.

SPECIFICATIONS AND THEOREMS

Assume that we are given n observations of the real variables X_1, $X_2, \cdots, X_k$, and

$$Y = \psi + \epsilon \tag{1}$$

where each ϵ is a random selection from some universe such that, for every observation,

$$E[X_i\epsilon] = E[\psi\epsilon] = E[\epsilon] = 0, \quad (i = 1, 2, \cdots, k), \tag{2}$$

and where

$$\psi = E[Y], \tag{3}$$

the symbol E denoting expected value. Make no assumption as to relationship between $X_1, X_2, \cdots, X_k$, and ψ. Let

$$f = B_1X_1 + B_2X_2 + \cdots + B_kX_k \tag{4}$$

be the general expression for one of the n values of the linear regression function of the X's, fitted to the n observed values of Y by the method of least squares; that is, let the regression coefficients B_1, $B_2, \cdots, B_k$ be computed by solving the k normal equations obtained by setting the partial derivative of

$$\sum (B_1X_1 + B_2X_2 + \cdots + B_kX_k - Y)^2$$

with respect to each B equal to zero ($\sum$ denoting summation over all n observations). Assume that the solution is unique; and that it is possible to construct the k orthogonal linear functions

$$\left.\begin{aligned}
x_1 &= X_1, \\
x_2 &= C_{21}X_1 + X_2, \\
x_3 &= C_{31}X_1 + C_{32}X_2 + X_3, \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
x_k &= C_{k1}X_1 + C_{k2}X_2 + \cdots + C_{k,k-1}X_{k-1} + X_k,
\end{aligned}\right\} \tag{5}$$

with coefficients $C_{21}, C_{31}, \cdots, C_{k,k-1}$ such that

$$\left.\begin{array}{r} \sum x_i x_j = 0, \\ \sum x_i^2 \neq 0, \end{array}\right\} \tag{6}$$

for $i=1, 2, \cdots, k$; $j=2, 3, \cdots, k$; $j\neq i$. Let

$$\phi = E[f] \tag{7}$$

at a particular observation, for the set of n values of $\psi, X_1, X_2, \cdots, X_k$. Let

$$\sigma^2 = E[\epsilon^2], \tag{8}$$

and assume σ^2 to be finite. Then the main theorems to be derived may be written

$$E[\sum (f - \psi)^2] = \sum (\phi - \psi)^2 + k\sigma^2 \tag{A}$$

and

$$E[\sum (f - Y)^2] = \sum (\phi - \psi)^2 + (n - k)\sigma^2. \tag{B}$$

Of more practical importance are the corollaries

$$\sigma^2 \leqq E[\sum (f - Y)^2/(n - k)], \tag{I}$$

$$\frac{1}{n} \sum (\phi - \psi)^2 = \frac{1}{n} E[\sum (f - Y)^2] - (n - k)\sigma^2/n, \tag{II}$$

$$E\left[\frac{1}{n} \sum (f - \psi)^2\right] = \frac{1}{n} E[\sum (f - Y)^2] - (n - 2k)\sigma^2/n, \tag{III}$$

and

$$E\left[\frac{1}{n} \sum (f - \psi)^2\right] = \frac{1}{n} \sum \psi^2 + \text{the sum of } k \text{ terms of the form } \frac{1}{n}\left\{\sigma^2 - (\sum x_i\psi)^2/\sum x_i^2\right\}; \tag{IV}$$

also,

$$\delta^2 = \frac{1}{n} E[\sum (f - Y)^2] + 2k\sigma^2/n \tag{V}$$

and

$$\delta^2 = E\left[\frac{1}{n} \sum (f - \psi)^2\right] + \sigma^2 \tag{VI}$$

where δ^2 is defined by the relationships

$$\delta^2 = \frac{1}{n} E\left[\sum (f - Y')^2\right] \tag{9}$$

and

$$Y' = \psi + \epsilon', \tag{10}$$

each ϵ' being a random selection from the same universe as the ϵ's, but independent of the ϵ's included in the Y's that were used in computing the B's in equation (4).

DERIVATION

Assume that the coefficients $b_1, b_2, \cdots, b_k$ in the regression function

$$b_1x_1 + b_2x_2 + \cdots + b_kx_k$$

have been computed by the method of least squares so as to minimize

$$\sum (b_1x_1 + b_2x_2 + \cdots + b_kx_k - Y)^2.$$

(In solving actual problems it is not necessary to make the transformation to orthogonal functions; but the further discussion is simplified if we assume that it has been made.) Then from (6) we see the normal equations take the form

$$\left.\begin{aligned} b_1 \sum x_1^2 &= \sum x_1 Y, \\ b_2 \sum x_2^2 &= \sum x_2 Y, \\ &\cdots\cdots\cdots \\ b_k \sum x_k^2 &= \sum x_k Y. \end{aligned}\right\} \tag{11}$$

It is known that under these conditions,

$$b_1x_1 + b_2x_2 + \cdots + b_kx_k \equiv f \tag{12}$$

where f is defined by equation (4).

We are now ready to derive the sampling variance of f. From equations (11) and (12), we have[2]

$$f = \frac{x_1 \sum x_1 Y}{\sum x_1^2} + \frac{x_2 \sum x_2 Y}{\sum x_2^2} + \cdots + \frac{x_k \sum x_k Y}{\sum x_k^2}. \tag{13}$$

Replacing Y with $\psi + \epsilon$, we obtain

[2] Compare Davis, H. T., *The Analysis of Economic Time Series*, 1941, p. 60, equation (3); Jones, H. L., "Fitting Polynomial Trends to Seasonal Data by the Method of Least Squares," *Journal of the American Statistical Association*, vol. 38 (1943), p. 453, equations (7) and (8).

$$f = \frac{x_1 \sum x_1\psi}{\sum x_1^2} + \frac{x_2 \sum x_2\psi}{\sum x_2^2} + \cdots + \frac{x_k \sum x_k\psi}{\sum x_k^2}$$
$$+ \frac{x_1 \sum x_1\epsilon}{\sum x_1^2} + \frac{x_2 \sum x_2\epsilon}{\sum x_2^2} + \cdots + \frac{x_k \sum x_k\epsilon}{\sum x_k^2}. \tag{14}$$

From (7), (14), (5), and (2), we have

$$\phi = \frac{x_1 \sum x_1\psi}{\sum x_1^2} + \frac{x_2 \sum x_2\psi}{\sum x_2^2} + \cdots + \frac{x_k \sum x_k\psi}{\sum x_k^2}. \tag{15}$$

At a particular observation the sampling error in f is, by definition, $f - E[f]$; that is, $f - \phi$. Let

$$\zeta = f - \phi. \tag{16}$$

Then from (16), (14), and (15),

$$\zeta = \frac{x_1 \sum x_1\epsilon}{\sum x_1^2} + \frac{x_2 \sum x_2\epsilon}{\sum x_2^2} + \cdots + \frac{x_k \sum x_k\epsilon}{\sum x_k^2}. \tag{17}$$

From (5), (17), and (2),

$$E[x_i\zeta] = E[\zeta] = 0 \tag{18}$$

for every x_i and every observation. Also, from the specification that ϵ is selected at random, we have

$$E[\epsilon_r\epsilon_s] = 0 \tag{19}$$

where the subscripts r and s denote different observations. From (8) and (19), the expected value of the square of (17) may therefore be written

$$E[\zeta^2] = \frac{x_1^2 \sum x_1^2\sigma^2}{(\sum x_1^2)^2} + \frac{x_2^2 \sum x_2^2\sigma^2}{(\sum x_2^2)^2} + \cdots + \frac{x_k^2 \sum x_k^2\sigma^2}{(\sum x_k^2)^2}$$
$$+ \frac{2x_1x_2 \sum x_1x_2\sigma^2}{\sum x_1^2 \sum x_2^2} + \frac{2x_1x_3 \sum x_1x_3\sigma^2}{\sum x_1^2 \sum x_3^2} + \cdots$$
$$+ \frac{2x_{k-1}x_k \sum x_{k-1}x_k\sigma^2}{\sum x_{k-1}^2 \sum x_k^2},$$

or, from the orthogonal property of the x's, equations (6),

$$E[\zeta^2] = \sigma^2 \left\{ \frac{x_1^2}{\sum x_1^2} + \frac{x_2^2}{\sum x_2^2} + \cdots + \frac{x_k^2}{\sum x_k^2} \right\}. \tag{20}$$

This is the sampling variance of f at a particular observation. Summing over all n observations yields

$$E[\sum \zeta^2] = k\sigma^2. \tag{21}$$

We next derive the sampling covariance of Y and f. This may be defined by the equation

$$\pi[Yf] = E[Yf] - \{E[Y]\}\{E[f]\}$$

where π denotes sampling covariance. In this case, by substituting the equivalent values of Y, f, $E[Y]$, and $E[f]$ as given by (1), (16), (3), and (7), and by using (17), (5), (2), and (15) to eliminate terms equal to zero, we obtain

$$\pi[Yf] = E[\psi\zeta + \epsilon\phi + \epsilon\zeta] = E[\epsilon\zeta]. \tag{22}$$

Multiplying equation (17) by ϵ, we obtain

$$\epsilon\zeta = \frac{x_1\epsilon \sum x_1\epsilon}{\sum x_1^2} + \frac{x_2\epsilon \sum x_2\epsilon}{\sum x_2^2} + \cdots + \frac{x_k\epsilon \sum x_k\epsilon}{\sum x_k^2},$$

whence, from (8), (19), and (20),

$$E[\epsilon\zeta] = \sigma^2\left\{\frac{x_1^2}{\sum x_1^2} + \frac{x_2^2}{\sum x_2^2} + \cdots + \frac{x_k^2}{\sum x_k^2}\right\} = E[\zeta^2]; \tag{23}$$

that is, at each observation, the sampling covariance of Y and f is the same as the sampling variance of f. We have immediately

$$E[\sum \epsilon\zeta] = E[\sum \zeta^2] = k\sigma^2. \tag{24}$$

From (16), the expected value of the sum of the squared errors in f, measured from ψ, may be written

$$E[\sum (f - \psi)^2] = E[\sum (\phi - \psi + \zeta)^2]$$
$$= E[\sum (\phi - \psi)^2 + \sum \zeta^2 + 2\sum \zeta(\phi - \psi)].$$

From (21), (17), (15), (18), and (2), this equation becomes

$$E[\sum (f - \psi)^2] = \sum (\phi - \psi)^2 + k\sigma^2, \tag{25}$$

which completes the derivation of Theorem (A).

Expanding the expression for the expected value of the sum of the squared residuals, we write

$$E[\sum (f-Y)^2] = E[\sum (f-\psi-\epsilon)^2]$$
$$= E[\sum (f-\psi)^2 + \sum \epsilon^2 - 2\sum \epsilon(\phi-\psi) - 2\sum \epsilon\zeta]. \tag{26}$$

From (15), (5), and (2),

$$E\left[\sum \epsilon(\phi - \psi)\right] = 0. \tag{27}$$

We have, therefore, from the combination of (26) with (25), (8), (27), and (24),

$$E\left[\sum (f - Y)^2\right] = \sum (\phi - \psi)^2 + (n - k)\sigma^2, \tag{28}$$

which completes the derivation of Theorem (B).

Corollaries (I), (II), and (III) are easily derived from Theorems (A) and (B).

The bias of the regression function, f, at a particular observation is defined to be the difference $\phi - \psi$. Its mean square for all n observations is given by Corollary (II). Some additional properties will now be derived. From (15) and (6) we have

$$\sum \phi^2 = \frac{(\sum x_1\psi)^2}{\sum x_1^2} + \frac{(\sum x_2\psi)^2}{\sum x_2^2} + \cdots + \frac{(\sum x_k\psi)^2}{\sum x_k^2} = \sum \phi\psi, \tag{29}$$

whence

$$\sum (\phi - \psi)^2 = \sum \psi^2 - \frac{(\sum x_1\psi)^2}{\sum x_1^2} - \frac{(\sum x_2\psi)^2}{\sum x_2^2} - \cdots - \frac{(\sum x_k\psi)^2}{\sum x_k^2}. \tag{30}$$

Combining (30) with Theorem (A), and dividing by n, leads to Corollary (IV).

It is to be noted that if a set of orthogonal functions $x_1, x_2, \cdots, x_i$, defined by equations (5) and (6), has been computed for n observations of $X_1, X_2, \cdots, X_i$, the addition of a new variable X_{i+1} and then transforming to a new set of $x_1, x_2, \cdots, x_{i+1}$, satisfying (5) and (6), requires the computation of but one new orthogonal function, x_{i+1}, the functions $x_1, x_2, \cdots, x_i$ remaining the same as before. From this fact and equation (30), it follows that adding a variable X_{i+1} to an original set of variables $X_1, X_2, \cdots, X_i$ and then recomputing a least square linear regression function of the variables can not possibly increase the mean square bias, since each term following $\sum\psi^2$ in equation (30) must be negative or zero.

A more general statement follows from a comparison of equations (13) and (15). Since f and ϕ are seen to be of the same functional form, it is obvious that among all functions of the linear form

$$\Phi = \beta_1X_1 + \beta_2X_2 + \cdots + \beta_kX_k \tag{31}$$

the particular one that minimizes $(1/n)\sum(\Phi-\psi)^2$, is ϕ as defined by (15). In other words, if

$$\phi' = \beta_1'X_1 + \beta_2'X_2 + \cdots + \beta_k'X_k \tag{32}$$

is a linear function of $X_1, X_2, \cdots, X_k$, only, such that

$$\phi' \neq \phi \tag{33}$$

for at least one observation, then

$$\sum (\phi - \psi)^2 < \sum (\phi' - \psi)^2. \tag{34}$$

(A linear function with a constant term is to be regarded as the special case where $X_1 \equiv 1$; otherwise, it might be an exception.)

To derive Corollary (V), we first expand $E[\sum(f-Y')^2]$, obtaining

$$\begin{aligned} E[\sum (f-Y')^2] &= E[\sum (\phi+\zeta-\psi-\epsilon')^2] \\ &= E[\sum (\phi-\psi)^2 + \sum \zeta^2 + \sum (\epsilon')^2 + 2\sum \zeta(\phi-\psi) \\ &\quad -2\sum \epsilon'(\phi-\psi) - 2\sum \zeta\epsilon']. \end{aligned} \tag{35}$$

From the specifications of ϵ' following equation (10), we write

$$E[\zeta\epsilon'] = E[\epsilon'] = 0, \tag{36}$$

and

$$E[(\epsilon')^2] = \sigma^2. \tag{37}$$

Equation (35) therefore reduces to

$$E[\sum (f - Y')^2] = \sum (\phi - \psi)^2 + (n + k)\sigma^2. \tag{38}$$

Combining (38) with Theorem (B) yields

$$E[\sum (f - Y')^2] = E[\sum (f - Y)^2] + 2k\sigma^2. \tag{39}$$

Dividing by n, and combining with (9), completes the derivation of Corollary (V). Combining (III) and (V) leads to Corollary (VI).

PRACTICAL APPLICATIONS

The theorems and corollaries in the preceding sections were derived for the general case where the relationship between the independent variables and the expected value of the dependent variable is not specified. This case is broad enough to cover situations where the linear regression function neglects some of the independent variables, including terms of the second and higher degrees in the variables that do appear in this function. More simply, we can always think of the

differences between the expected values of the dependent variable and the expected values of a linear regression function, when such differences exist, as constituting a set of values of a single neglected variable. In the following paragraphs, the application of the theorems and corollaries to practical regression problems will be considered for the case where one or more variables are neglected.

Suppose several independent variables are being considered, and it is desired to investigate various combinations for predicting the dependent variable or estimating its expected value. Let m be the number of such independent variables (including $X_1 \equiv 1$ if the regression function is to include a constant term, B_1), and let n be the number of observations. (It is desirable that $n-m$ be large.) We may first compute the sum of the squared residuals for the least square linear regression function of all m variables and divide by $n-m$. Let us designate the quotient s^2. Then from Corollary (I) we know that the expected value of s^2 is equal to or greater than σ^2, the sampling variance of the dependent variable. In this sense, s^2 is a conservative estimate of σ^2.

We see from inequality (34) that the m-variable function has a mean square bias less than, or equal to, the mean square bias of any other function of the same m variables that is of the same linear form—for example, any such function where one or more of the variables are, in effect, neglected by setting their coefficients equal to zero. Suppose we assume the m-variable function to be unbiased. We can then use Corollary (II) to estimate the mean square bias of any other linear regression function, of any k variables. We compute the sum of the squared residuals for this k-variable function, and subtract $(n-k)s^2$ where s^2 is the estimate of σ^2 previously obtained by employing all m variables. The remainder, divided by n, is the estimated mean square bias. If the quotient should turn out to be negative for a regression function which employs no variable not included among the independent variables used in computing s^2, it should usually be ascribed to chance and not interpreted as evidence that the k-variable function has a smaller mean square bias than the m-variable function, as inequality (34) shows this latter interpretation to be incorrect under the assumed conditions. Note that if the estimate of σ^2 is too large, the usual tendency, the mean square bias will tend to be underestimated for all the regression functions, the expected value of the error being proportional to $(n-k)$. In many comparisons, however, these estimates may nevertheless be indicative of the rank of different regression functions with respect to bias.

To estimate the mean square error, as measured from the expected values of the dependent variable, Corollary (III) suggests that the mean square residual be reduced by $(n-2k)s^2/n$, where s^2 is computed as outlined above. The special case where $k=\frac{1}{2}n$ is interesting, for then the mean square error and the mean square residual have the same expected value. If the m-variable function used to compute s^2 is biased, however, the estimates of the mean square error will tend to be too low for $k<\frac{1}{2}n$, and too high for $k>\frac{1}{2}n$. For the m-variable function, the combination of Corollaries (I) and (III) suggests estimating the mean square error by multiplying the mean square residual by $m/(n-m)$. As before, the resulting estimate will be unbiased if the regression function is unbiased at every observation; otherwise, the estimate will tend to be too low for $m<\frac{1}{2}n$, and too high for $m>\frac{1}{2}n$.

In many problems of estimating a variable, it is useful to estimate the quantity δ^2 as defined by equation (9). It is the expected value of the square of the standard error in estimating a new set of values of the dependent variable where these values are assumed to be constructed by replacing the random elements in the original observations by new random selections from the same universe. In general, δ^2 is equal to the sum of (1) the expected value of the mean square of the differences between the estimates and the observed values of the dependent variable, and (2) twice the mean sampling covariance of these observed values and the estimates.[3]

To estimate δ^2 for a least square linear regression function, Corollaries (I) and (V) suggest that the mean square residual be increased by $2ks^2/n$ where s^2 is computed from the m-variable regression function previously defined. In making a choice among several regression functions, δ^2 may be estimated for each function, using the same s^2, to determine the function for which the estimate of δ^2 is smallest. For the m-variable function, we may estimate δ^2 by multiplying the mean square residual by $(n+m)/(n-m)$. Corollary (VI) shows that δ^2 differs from the expected value of the mean square error by the constant σ^2. It is therefore immaterial whether one or the other of these two criteria is used in choosing a regression function, provided the same estimate of σ^2 is used. In either case, if σ^2 is overestimated, the selection will be biased in favor of the regression function with the smaller number of variables. However, estimates of δ^2 have the important property, not possessed by estimates of the mean square error, that their expected values are always equal to or greater than the quantities that are being estimated, for a random selection of regression function under the assumed condi-

[3] It is difficult to think of a suitable name for δ. Perhaps *standard residual* would do.

tions as to the random component of the dependent variable. In this sense they are conservative estimates. When the estimates themselves are used to determine the choice of function, however, there may be a tendency to underestimate δ^2 for the function chosen.

There is an interesting relationship between the estimated value of δ^2, as a criterion for deciding which variables to neglect, and the usual tests of significance in multiple correlation and the analysis of variance where the null hypothesis states that the partial regression coefficient of the expected values of the dependent variable on the observed values of an independent variable is zero. The δ^2 criterion rejects the independent variable—that is, sets its coefficient in the estimating function equal to zero—unless t^2, or F, is greater than 2, where t and F refer to "Student's" t and Snedecor's F, respectively. To make this clear, let us suppose that we are given n observations of the dependent variable Y and the independent variables $X_1, X_2, \cdots, X_k, \cdots, X_l, \cdots, X_m$, $k<l\leqq m$, and that we have computed the linear regression functions f_k, f_l, and f_m from the first k, the first l, and all m variables, respectively. Then if d_k^2 and d_l^2 are estimates of δ^2 for f_k and f_l, computed from the formulas

$$s^2 = \sum (f_m - Y)^2/(n - m), \tag{43}$$

$$d_k^2 = \frac{1}{n}\left[\sum (f_k - Y)^2 + 2ks^2\right], \tag{44}$$

and

$$d_{oll}^2 = \frac{1}{n}\left[\sum (f_2 - Y)^2 + 2ls^2\right], \tag{45}$$

and if we compute

$$F = \frac{\left[\sum (f_k - Y)^2 - \sum (f_2 - Y)^2\right]/(l - k)}{\sum (f_m - Y)^2/(n - m)}, \tag{46}$$

it can easily be shown that $d_l^2<d_k^2$ if and only if $F>2$. The critical value of F (or t^2) corresponds to fairly large probabilities with respect to the null hypothesis referred to at the beginning of this paragraph.

The proper choice of criterion depends, of course, upon the purpose in applying it. If it is important not to include an independent variable in the estimating function unless there is strong evidence that the partial correlation of the variable with the expected values of the dependent variable is different from zero, the t or F test may be applied in the search for such evidence. On the other hand, if it is important to have unbiased estimates, then it must be kept in mind that a value of t^2 or F greater than one supplies some evidence that setting the tested

coefficient or coefficients equal to zero will result in biased estimates; and that there is no test, based on sampling theory, which supplies satisfactory evidence that the "true" partial regression coefficient is exactly zero (except in special situations, as when admissible hypotheses state the coefficient to be zero or one). In deciding which variables to neglect, the δ^2 criterion should not be used in attempting to answer the question as to whether the "true" regression coefficient is different from zero, but only in trying to find the function that results in estimates with the smallest mean square error.

Corollary (IV) throws light on the problem of choosing a criterion. It shows that adding one more variable, X_i, to a linear regression function reduces the expected value of its mean square error if and only if

$$(\sum x_i\psi)^2/\sum x_i^2 > \sigma^2, \tag{47}$$

the left-hand member of this inequality, divided by n, representing the reduction in the mean square bias, and the right-hand member, divided by n, representing the increase in the mean sampling variance. In testing a single variable, X_i, the objective in using the δ^2 criterion is to retain this variable in the regression function when inequality (47) is true, and to neglect X_i when (47) is false. An underlying premise is that the risk of accepting (47) when it is false is of about the same importance as the risk of rejecting (47) when it is true. On the other hand, the objective in using the F or t test is to neglect the variable X_i in the regression function when

$$(\sum x_i\psi)^2/\sum x_i^2 = \sum x_i\psi = 0, \tag{48}$$

and to retain the variable otherwise. The common use of this test with a small probability as the basis for making decisions rests on the premise that the risk of rejecting (48) when it is true is very important, while the risk of accepting (48) when it is false is comparatively unimportant. On account of the difference in the relative importance attached to the two kinds of risks, the F test sometimes leads to the nonrejection of (48) where the δ^2 criterion leads to the acceptance of (47). As usually employed, the F test frequently tells us to retain a variable in the regression function, but never tells us to neglect one. The δ^2 criterion does both, the decision to retain a variable depending on whether the gain in reducing the mean square bias appears to outweigh the increase in the mean sampling variance.

In fitting empirical functions to a set of observations there is no assurance, of course, that δ^2 or any other criterion will be satisfactory

for unobserved points outside or inside the region of the observations, or even for recurrences of the observed combinations of values of the independent variables if some variable not considered affects the dependent variable. There are also problems where our assumptions as to the random component will not be sufficiently realistic. Only practical experience can determine which assumptions and what criteria are most useful for any given type of problem in a particular field of research.

Linear restrictions. In computing partial regression coefficients B_1, $B_2, \cdots, B_m$, it is sometimes desirable to impose the condition that the computed values satisfy c consistent independent equations of the form

$$\alpha_{i1}B_1 + \alpha_{i2}B_2 + \cdots + \alpha_{im}B_m = \gamma_i, \quad (i = 1, 2, \cdots, c), \quad (49)$$

where the α's and γ's are specified constants. A special case is that where these equations are all of the simple form

$$B_i = 0. \quad (50)$$

Imposing the condition that the computed values of the B's satisfy c equations of this simple form is equivalent, of course, to neglecting c independent variables in computing the regression function.

In general, it can be shown[4] that requiring the solution for the regression coefficients to satisfy c equations of form (49) is equivalent to making a linear transformation of the variables Y, X_1, $X_2, \cdots, X_m$ and then neglecting c transformed independent variables. It can also be shown that if, for every observation, equations (2) and (3) are true for ψ, ϵ, and every X_i in terms of the original variables, they are also true in terms of the transformed variables, σ^2 and each ϵ remaining unchanged. Hence, if the transformed variables can be orthogonalized, all our theorems and corollaries hold, provided k be replaced with $m-c$.

[4] See Wilks, S. S., *Mathematical Statistics*, 1943, pp. 171–172. There it is specified that the random component in the dependent variable is normally distributed about a linear function of the independent variables. The discussion of the transformation is sufficiently general, however, to cover the case considered in this article where neither the functional relationship between the variables nor the form of the distribution of the random component is specified.

Incidentally, this discussion settles a question raised elsewhere by the present writer as to whether a least square solution results in the unbiased estimate with minimum sampling variance when the solution is subject to a linear restriction on the parameters. See reference in footnote 3 *supra*, p. 459.

THE ANALYSIS OF LATIN SQUARES WHEN SOME OBSERVATIONS ARE MISSING*

D. B. De Lury
Virginia Polytechnic Institute

The discussion of the missing-value problem is given explicitly for a biological assay which employs a 4×4 latin square in several replications. However, the methods are easily adapted to any latin square and to various other designs as well.

Methods of analysis, when some observations are missing, are discussed for the following cases.

(1) One or more "single" observations are missing.
(2) Several columns are missing.
(3) One column is missing.
(4) Two columns are missing.
(5) One column and one or more single observations are missing.

Introduction. Devices which can be used to simplify the statistical analysis when some of the observations are missing from a latin square have been discussed by Yates [1] and by Yates and Hale [2]. These discussions cover most of the cases of practical importance, but in certain situations, for example, when a latin square of fixed size is in routine use, the methods and results given in these papers can be reorganized so as to increase considerably the simplicity of the formulae and of the arithmetic required for their application. The latin square arrangement suggested by Bliss and Marks [4] for use in the assay of insulin illustrates such a situation. This paper discusses the analysis, when some observations are missing, of this specific design, which is described in section 1. Modifications appropriate to other designs should be fairly obvious.

1. This assay is designed to compare the potencies of two preparations, which may conveniently be called the "standard" and the "unknown." Each preparation is administered at two dosage levels, which usually are made the same for both preparations under the assumption that the potencies are equal. Each experimental animal is given each of the four doses on four different days. The observations on four animals are arranged in a latin square, "rows" corresponding to days, "columns" to animals, and "treatments" to doses. This latin square is repeated

* A paper presented at the 105th Annual Meeting of the American Statistical Association, to a joint session of the Biometrics Section and the Institute of Mathematical Statistics, Cleveland, January 25, 1946.

several times, using different sets of animals, each set with its own random arrangement of doses, but administering the doses on the same days. Thus if r replications are used, the assay involves $4r$ animals and yields $16r$ observations.

Assuming that the effects of rows, columns and treatments combine additively, the analysis of variance table reads as follows when no observation is missing.

Sources of variation	*Degrees of freedom*
columns (animals)	$4r-1$
rows (days)	3
treatments (doses)	3
residual (error)	$12r-6$

The three degrees of freedom exhibiting differences among treatments are subdivided into

1 for the distance between the two dose-response lines, (D)
1 for the common slope of the two lines, (B)
1 for the departure from linearity of the dose-response curve over the range of the test (C).

Since the response is linearly related to the logarithm of the dose, the difference between the logarithms of the potencies of the two preparations is given by $I\,D/B$, where I represents the difference between the logarithms of the two doses. ([4] page 188.) Validity of the assay requires that B differ significantly from zero and that C be not significantly different from zero.

These calculations may be illustrated on the data of Table I, which gives the observations, in milligrams of glucose per 100 cc. of blood, obtained in a rabbit assay of insulin. The latin square arrangement, with 4 replications, was used. The treatments, assuming the same potency for unknown and standard, were:

treatment 1—0.30 units of the unknown preparation,
" 2—0.60 units of the unknown preparation,
" 3—0.30 units of the standard preparation,
" 4—0.60 units of the standard preparation.

The number of the treatment is recorded under each observation.

The analysis of variance table is made up using the ordinary rules for such computations. The sum of squares arising from differences among columns is given by $[(210)^2+(246)^2+\cdots+(288)^2]/4-(3782)^2/64$, and similar calculations furnish the sums of squares for rows and treat-

ments. The residual sum of squares is obtained by subtracting from the "total" sum of squares, those for columns, rows, and treatments.

	d.f.	s.s.	m.s.
columns	15	10,327	
rows	3	937	
treatments	3	3,486	
residual	42	2,589	61.65
total	63	17,339	

No importance is attached to testing for significance the treatment sum of squares, but since the validity of the assay depends on the significance or non-significance of the quantities B and C, these must be tested. It is convenient to calculate B, C and D according to the following scheme. ([4], page 184.)

treatment number	1	2	3	4		
treatment total	1078	845	1047	812	divisors	
D	$+1$	$+1$	-1	-1	$4\sqrt{r}=8$	$+8.00$
B	-1	$+1$	-1	$+1$	$4\sqrt{r}=8$	-58.50
C	-1	$+1$	$+1$	-1	$4\sqrt{r}=8$	$+0.25$

This array implies that $D=[(1078)(+1)+(854)(+1)+(1047)(-1)+(812)(-1)]/8=8.00$, etc.

The value of B may be tested either by calculating $t=|B|/s$, where s^2 is the residual mean square, and entering the t-table with 42 d.f., or by calculating $F=B^2/s^2$ and entering the F-table with $n_1=1$, $n_2=42$. Here the value of B is highly significant ($t=7.45$, $P<10^{-9}$). The value of C may be tested in exactly the same way. In this case it is obvious that C does not differ significantly from zero.

Since these tests give no reason for doubting the validity of the assumptions underlying the analysis of the assay, the logarithm of the ratio of the potencies may be computed as

$$M = (0.30103)(8.00)/(-58.50) = -.0412,$$

with an estimated standard error given by

$$s_M = \frac{sI\sqrt{B^2 + D^2}}{B^2} = 0.0320. \quad ([4], \text{ formula } (4).)$$

2. The simple analysis outlined above breaks down when an observation is missing and constants must be fitted to separate out the contributions of rows, columns and treatments. In order to set up the problem in symbols, let y_{ijk} represent the observation in row i and

column j, where $k=k\ (i, j)$ denotes the treatment in row i and column j provided by the latin square arrangement. Under the same assumptions that were made in the analysis given above, the equation representing the dependence of y on its row, column and treatment may be written

$$(2.1) \qquad Y_{ijk} = m + r_i + c_j + t_k, \quad \text{with} \quad \sum r_i = \sum c_j = \sum t_k = 0.$$

The r's, c's and t's represent estimates of the deviations, from the general mean, of the average contributions to the observed values of rows, columns, and treatments. m is the estimate of the general mean. Y_{ijk} is the estimate of the mean response to treatment k in row i and column j. In order to bring it into conformity with the standard form of regression equation, this equation may be written

$$(2.2) \qquad Y_{ijk} = m + \sum_{\alpha} r_\alpha \delta_{\alpha i} + \sum_{\beta} c_\beta \delta_{\beta j} + \sum_{\gamma} t_\gamma \delta_{\gamma k},$$

where $\delta_{pq}=1$ or 0 according as $p=q$ or $p\neq q$, and the summations extend over all rows, columns and treatments remaining in the assay. (This form of the equation is needed in section 5.) The δ's are the independent variables and the r's, c's, t's, and m are regression coefficients, chosen to minimize $S(y-Y)^2$, where S denotes summation over all the observations. The minimizing of this sum, subject to the restrictions $\sum r_i=0$ etc., may be carried out by introducing Lagrange multipliers and finding the unrestricted minimum of

$$\phi = S(y - Y)^2 + 2\lambda \sum r_i + 2\mu \sum c_j + 2\nu \sum t_k.$$

The derivatives of ϕ with respect to m, the r's, c's and t's, and λ, μ and ν, equated to zero, provide a set of normal equations which determine the values of the regression coefficients.

When no observation is missing these equations are very simple. Using

G	to denote the sum of all the observations,	
R_i	" " " " " " " "	in row i
C_j	" " " " " " " "	in column j,
T_k	" " " " " " " "	on treatment k,

the normal equations are

$$(2.3) \qquad \begin{aligned} 16rm &= G, & & \\ 4rm + 4rr_i + \lambda &= R_i, & i &= 1, 2, 3, 4, \\ 4m + 4c_j + \mu &= C_j, & j &= 1, 2, \cdots 4r, \\ 4rm + 4rt_k + \nu &= T_k, & k &= 1, 2, 3, 4. \end{aligned}$$

The solutions follow easily.

$$(2.4)\qquad \begin{aligned} m &= G/16r, \qquad r_i = R_i/4r - G/16r, \qquad c_j = C_j/4 - G/16r,\\ t_k &= T_k/4r - G/16r, \qquad \lambda = \mu = \nu = 0. \end{aligned}$$

The residual sum of squares, obtained by applying a standard least-squares formula, is given by

$$(2.5)\qquad S(y - Y)^2 = Sy^2 - mG - \sum r_iR_i - \sum c_jC_j - \sum t_kT_k.$$

Now $Sy^2 - mG = Sy^2 - G^2/16r$ is the "total" sum of squares of the analysis of variance, and $\sum r_iR_i = \sum R_i(R_i/4r - G/16r) = (\sum R_i^2)/4r - G^2/16r)$ is the sum of squares properly ascribed to variation among rows, undisturbed by differences among columns and among treatments. Similarly,

$$\sum c_jC_j = (\sum C_j^2)/4 - G^2/16r \text{ is the "column" sum of squares,}$$

$$\sum t_kT_k = (\sum T_k^2)/4r - G^2/16r \text{ is the "treatment" sum of squares.}$$

These sums of squares are the same as those entered in the analysis of variance table of section 1. It is to be remarked also that the combinations of observations designated B, C, D, are linear functions of t_1, t_2, t_3, t_4. For example, $B = \sqrt{r}(-t_1 + t_2 - t_3 + t_4)$.

3. When some observations are missing, the setting-up and solving of the normal equations are in general very laborious. However, in some cases, the equations are simple or can be made so by employing suitable devices; in others, the normal equations may be avoided entirely by using "missing-plot" formulae. The purpose of this paper is to explore these two situations, providing appropriate formulae when possible and illustrating devices which may be used when a solution by means of formulae is not practicable. No particular novelty is claimed for most of this discussion. It represents simply a collection of methods, most of them well known, for application to a specific problem.

The cases to be considered are the following:

(1) one or more single observations are missing;

(2) several columns are missing;

(3) one column is missing;

(4) two columns are missing;

(5) one column and one or more single observations are missing.

Cases (1), (3) and (5) may be treated by means of formulae; cases (2) and (4) require the solution of a set of normal equations. The more general situation in which several columns and one or more single observations are missing is not discussed, since no simple formulae are available in such cases.

4. Case (1). When one or more single observations are missing, the solving of normal equations may be avoided by applying the argument given by Yates [3], which, in brief, runs as follows. Suppose that the normal equations have been solved for the values of m, the r's, c's and t's. Then $Y_{ijk}=m+r_i+c_j+t_k$ is determined so that $S(y-Y)^2$ is a minimum, subject of course to the restrictions $\sum r_i=0$, etc. Suppose also that the Y-values corresponding to the missing observations are used to fill in the gaps in the table of observations. Now let $Y'_{ijk}=m'+r'_i+c'_j+t'_k$ be a regression equation fitted to the numbers in this completed table. This amounts to choosing Y' to minimize $S(y-Y')^2+\sum(Y-Y')^2$, where S represents, as before, a summation over the actual observations and $\sum$ here indicates summation over the values which were filled in from the regression equation. It is clear that this sum is minimized by putting $Y'=Y$, that is, the regression equation fitted to the completed table is identical with that fitted to the actual observations alone. The fitting to the completed set of observations is immediate, since the solutions obtained in section 2 now apply. It follows that when the analysis shown in section 1 is run on the completed table of observations, the error term and the components B, C, D are strictly correct. Hence if a way can be found to determine the Y-values to substitute for the missing observations, *without first fitting the regression equation*, a simple and direct method of analysis is provided. This is easily done.

Suppose that a single observation is missing, say in row l and column m. Let the symbol y_{lmn} be written for the number which ultimately will be substituted in this position. When a regression equation $Y_{ijk}=m+r_i+c_j+t_k$ is fitted to this "completed" table, the coefficients, $m=G/16r$, $r_i=R_i/4r-G/16r$, etc. (from section 2) involve the symbol y_{lmn}. Substitution for m, r_i, c_j, t_k in the regression equation yields the relation

$$Y_{ijk} = R_i/4r + C_j/4 + T_k/4r - 2G/16r.$$

Now the value of Y_{lmn} is obtained by equating y_{lmn} to Y_{lmn}. This forms a linear equation in y_{lmn} whose solution is

$$y_{lmn} = \frac{4rC'_m + 4(R'_l + T'_n) - 2G'}{12r - 6}. \tag{4.1}$$

The primes on the symbols indicate that these totals, all of which involve the cell of the table from which the observation is missing, are sums only of the actual observations. Thus C'_m, for example, represents the sum of the three observations in the column from which the obser-

vation is missing. When $r=1$, this formula reduces to that given by Yates ([3], page 132) for a single 4×4 latin square.

Formulae for obtaining simultaneously values to replace several missing observations may be derived in the same manner. However it is simpler to use an iterative process based on the single-value formula. This procedure is described in a number of places ([3] page 133, [5] page 263).

The sum of squares of residuals may be computed according to formula (2.5), that is, according to the rules used when no observation is missing, with the totals made up to include the numbers substituted for the missing values. The number of degrees of freedom associated with this sum of squares is $12r-6$ less the number of missing observations.

The sum $\sum r_i R_i$ cannot now be identified with variation arising solely from differences among rows, owing to correlations among the row, column and treatment constants. Likewise, the sums $\sum c_j C_j$ and $\sum t_k T_k$ do not represent variation among columns only and among treatments only. However, the effects of these correlations are small enough to be ignored, unless the number of missing observations is large ([3] page 138). The effect of a missing observation on the variances of B, C, and D is discussed briefly in section 8.

5. Case (2). When the missing observations make up complete columns it is, in general, necessary to solve a set of normal equations and the only question requiring discussion concerns the form in which this calculation is best carried out. Normal equations may be set up and solved for the constants r_i, c_j, t_k as in section 2 and the combinations of the t's which give the values of B, C, D may be calculated. However, it seems simpler to write the regression equation in such a form as to yield these combinations directly. This may be done by rearranging equation (2.2) into the identically equivalent form

$$(5.1)\qquad Y_{ijk} = m + \sum_{\alpha=1}^{3} r_\alpha^* u_{\alpha ijk} + \sum_{\beta=1}^{q-1} c_\beta^* v_{\beta ijk} + \sum_{\gamma=1}^{3} t_\gamma^* w_{\gamma ijk}$$

where

$$u_{1ijk} = \delta_{1i} + \delta_{2i} - \delta_{3i} - \delta_{4i} \qquad w_{1ijk} = \delta_{1k} + \delta_{2k} - \delta_{3k} - \delta_{4k}$$

$$u_{2ijk} = -\delta_{1i} + \delta_{2i} - \delta_{3i} + \delta_{4i} \qquad w_{2ijk} = -\delta_{1k} + \delta_{2k} - \delta_{3k} + \delta_{4k}$$

$$u_{3ijk} = -\delta_{1i} + \delta_{2i} + \delta_{3i} - \delta_{4i} \qquad w_{3ijk} = -\delta_{1k} + \delta_{2k} + \delta_{3k} - \delta_{4k}$$

$$v_{\beta ijk} = \delta_{1j} + \delta_{2j} + \cdots + \delta_{\beta j} - \beta\delta_{(\beta+1)j}, \qquad \beta = 1, 2, \cdots, (q-1),$$

and

$$r_1^* = (r_1 + r_2 - r_3 - r_4)/4 \qquad t_1^* = (t_1 + t_2 - t_3 - t_4)/4$$
$$r_2^* = (-r_1 + r_2 - r_3 + r_4)/4 \qquad t_2^* = (-t_1 + t_2 - t_3 + t_4)/4$$
$$r_3^* = (-r_1 + r_2 + r_3 - r_4)/4 \qquad t_3^* = (-t_1 + t_2 + t_3 - t_4)/4$$
$$c_\beta^* = (c_1 + c_2 + \cdots + c_\beta - \beta c_{\beta+1})/\beta(\beta+1), \quad = 1, 2, \cdots, (q-1).$$

The symbol q is used in these formulae to denote the number of columns (animals) remaining in the assay. There are no restraints on the values of the regression coefficients when the equation is written in this form.

The w's and t^*'s have been chosen to make t_1^*, t_2^*, t_3^* proportional to D, B, C. The choice of the u's and r^*'s is based chiefly on considerations of symmetry, since the values of the row constants are of no direct interest in this example. Likewise the column constants play no direct part in the evaluation of the assay and may be chosen in any way considered suitable. In this particular case, no simplicity is gained by replacing the original independent variables (the $\delta_{\beta j}$) and column constants (the c_j) by orthogonal combinations of them (such as the v's and c^*'s). However, this change of variables may well prove useful, in other situations of this kind, in simplifying a system of normal equations. Other orthogonal functions may, of course, be used, according to the needs of the particular situation.

Assuming that no observation is missing, apart from the missing columns, all sums and sums of products of the independent variables (the u's, v's and w's) vanish except those of form $S(u_{\alpha ijk}\ w_{\gamma ijk})$, $\alpha, \gamma = 1, 2, 3$. These sums may be written

$$\sum_{i=1}^{4} u_{\alpha ijk} \sum_{j=1}^{q} w_{\gamma ijk} = \sum_{i=1}^{4} a_{i\gamma} u_{\alpha ijk},$$

by letting $a_{i\gamma}$ represent the value of $\sum_{j=1}^{q} w_{\gamma ijk}$. Thus $a_{i\gamma}$ is simply the sum of the values taken by $w_{\gamma ijk}$ over row i. These numbers $a_{i\gamma}$ are always zero for the intact design, but if some columns are missing, their values depend on the arrangement of treatments in the missing columns. These values are always small positive or negative integers and are easily calculated. In all cases $\sum_{i=1}^{4} a_{i\gamma} = 0$, $\gamma = 1, 2, 3$, which provides a useful check on this calculation. When the values of the $a_{i\gamma}$ have been calculated, the sums $\sum_{i=1}^{4} a_{i\gamma} u_{\alpha ijk} = A_{\alpha\gamma}$ (say) may be evaluated. The other coefficients needed for the normal equations are

$Su^2_{\alpha ijk} = Sw^2_{\gamma ijk} = 4q, \quad Sv^2_{\beta ijk} = 4\beta(\beta + 1), \quad \beta = 1, 2, \cdots, (q - 1).$

The normal equations are

$$
\begin{aligned}
4qm &= G,\\
4\beta(\beta+1)c_\beta^* &= C_\beta^*, \quad \beta = 1, 2, \cdots, (q-1)\\
4qr_1^* + A_{11}t_1^* + A_{12}t_2^* + A_{13}t_3^* &= R_1^*\\
4qr_2^* + A_{21}t_1^* + A_{22}t_2^* + A_{23}t_3^* &= R_2^*\\
4qr_3^* + A_{31}t_1^* + A_{32}t_2^* + A_{33}t_3^* &= R_3^*\\
A_{11}r_1^* + A_{21}r_2^* + A_{31}r_3^* + 4qt_1^* &= T_1^*\\
A_{12}r_1^* + A_{22}r_2^* + A_{32}r_3^* + 4qt_2^* &= T_2^*\\
A_{13}r_1^* + A_{23}r_2^* + A_{33}r_3^* + 4qt_3^* &= T_3^*
\end{aligned}
$$

where $C_\beta^*, R_1^*, \cdots T_3^*$ stand for the following combinations of the observations.

$$
\begin{aligned}
&C_\beta^* = C_1 + C_2 + \cdots + C_\beta - \beta C_{\beta+1}, \qquad \beta = 1, 2, \cdots, (q-1)\\
&R_1^* = R_1 + R_2 - R_3 - R_4 \qquad T_1^* = T_1 + T_2 - T_3 - T_4\\
&R_2^* = -R_1 + R_2 - R_3 + R_4 \qquad T_2^* = -T_1 + T_2 - T_3 + T_4\\
&R_3^* = -R_1 + R_2 + R_3 - R_4 \qquad T_3^* = -T_1 + T_2 + T_3 - T_4.
\end{aligned}
$$

The residual sum of squares is calculated from the formula

$$
\begin{aligned}
S(y - Y)^2 = [Sy^2 - mG] - \left[\sum_{\beta=1}^{q-1} c_\beta^* C_\beta^*\right]\\
- \left[\sum_{\alpha=1}^{3} r_\alpha^* R_\alpha^* + \sum_{\gamma=1}^{3} t_\gamma^* T_\gamma^*\right].
\end{aligned} \tag{5.2}
$$

The term in the first square bracket is the "total" sum of squares of deviations about the mean, usually calculated according to the formula $Sy^2 - G^2/4q$; the second term is the sum of squares attributable to variation among columns and is best calculated from the equivalent formula $\sum_{j=1}^{q} C_j^2/4 - G^2/4q$; the third term derives its value from variation due to "rows and treatments." The sums $\sum r_\alpha^* R_\alpha^*$ and $\sum t_\gamma^* T_\gamma^*$ cannot be regarded separately as exhibiting variation among rows and among treatments. When such quantities are needed they may be obtained without difficulty, but they are not required here and the question will not be discussed. The only quantities needed for the analysis of the assay are $S(y-Y)^2$, the numbers t_1^*, t_2^*, t_3^*, and certain elements of the inverse matrix.

If the covariance of t_γ^* and $t_{\gamma'}^*$ is denoted $\sigma^2 c_{\gamma\gamma'}$, then $c_{\gamma\gamma'}$ is one of the elements of the inverse matrix. Its value may be obtained by putting unity for $T_{\gamma'}^*$ and zeros for all other R^*'s and T^*'s in the expression for t^*_γ, obtained from the solution of the normal equations. This rule applies also when $\gamma=\gamma'$, in which case $\sigma^2 c_{\gamma\gamma}$ is the variance of t_γ^*. The value of σ^2 is estimated by $s^2=S(y-Y)^2/(3q-6)$.

Any one of the t^*'s may be tested for significant departure from zero, either by calculating $t=(t_\gamma^*/s\sqrt{c_{\gamma\gamma}})$ with $3q-6$ degrees of freedom, or $F=t^2$ with $n_1=1$, $n_2=3q-6$.

The estimate of the logarithm of the ratio of the potencies is

$$M = It_1^*/t_2^* \tag{5.3}$$

with an estimated standard error given by

$$s_M = \frac{Is}{t_2^{*2}}\sqrt{c_{11}t_2^{*2} - 2c_{12}t_2^*t_1^* + c_{22}t_1^{*2}}, \tag{5.4}$$

which is simply the form taken in this case by the first order approximation to the standard error of a ratio.

The computation of the numbers $A_{\alpha\gamma}$ from their definition is not difficult and does not require an inordinate amount of time. However, this calculation may be simplified and systematized considerably by regarding the matrix $(A_{\alpha\gamma})$ as the product of two matrices as follows (shortening the symbol $u_{\alpha ijk}$ to $u_{\alpha i}$):

$$\begin{pmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ u_{21} & u_{22} & u_{23} & u_{24} \\ u_{31} & u_{32} & u_{33} & u_{34} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}.$$

The u-matrix is determined by the choice of the functions $u_{\alpha ijk}$ and is fixed when this choice has been made. For the u's chosen in this paper, the u-matrix is

$$\begin{pmatrix} 1 & 1 & -1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{pmatrix}.$$

The a-matrix depends on the form of the functions $w_{\gamma ijk}$ and on the arrangement of treatments in the missing columns. When this matrix is determined for a given case, the A-matrix is calculated by matrix multiplication.

The solution of the normal equations is in general fairly laborious, but in the cases which are likely to occur most frequently the solution is simple or can be avoided altogether. A discussion of these cases follows.

6. Case (3). When a single column is missing, no essential restriction is imposed by the assumption that treament 1 is missing from row 1, treatment 2 from row 2, etc., since this can always be arranged by renumbering the rows. The A-matrix for this order is

$$\begin{bmatrix} -4 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & -4 \end{bmatrix}.$$

The solutions of the resulting normal equations are

$$(6.1)\quad r_\alpha^* = (qR_\alpha^* + T_\alpha^*)/4(q^2 - 1),\ t_\gamma^* = (qT'_\gamma{}^* + R_\gamma^*)/4(q^2 - 1) \qquad \alpha, \gamma = 1, 2, 3,$$

and the elements $c_{\gamma\gamma'}$ of the inverse matrix are

$$(6.11)\qquad c_{11} = c_{22} = c_{33} = q/4(q^2 - 1), \qquad c_{12} = c_{23} = c_{31} = 0.$$

Substitution of these values for the r^*'s and t^*'s in $\sum r_\alpha^* R_\alpha^* + \sum t_\gamma^* T'_\gamma{}^*$ yields the formula

$$(6.2)\quad \left(\sum_{\alpha=1}^{3} R_\alpha^{*2} + \sum_{\gamma=1}^{3} T_\gamma^{*2} \right) \Big/ 4(q+1) + \sum_{\alpha=1}^{3} (R_\alpha^* + T'_\alpha{}^*)^2/4(q^2 - 1)$$

for the reduction in the sum of squares due to rows and treatments.

Since this formula does not require the values of the r^*'s, these numbers would not ordinarily be computed.

It may be remarked that the sum of squares due to treatments alone (eliminating rows) is given by $\dfrac{4(q^2-1)}{q} \sum t_\gamma^{*2}$, since the t^*'s are mutually uncorrelated. Likewise the sum of squares due to rows (eliminating treatments) is $\dfrac{4(q^2-1)}{q} \sum r_\alpha^{*2}$. These facts are not needed for the analysis of the assay, however.

The formulae deduced in this section may be applied whenever only one column is missing. In this way the solution of normal equations is avoided and indeed the analysis is no more difficult and requires only slightly more computation than that of the intact design.

7. Case (4). When two columns are missing the solving of the normal equations cannot conveniently be avoided, but the determination of the A-matrix can be made very simple, and the equations are not particularly difficult to solve.

Let the symbol

$$\begin{bmatrix} (p_1, p_2) \\ (q_1, q_2) \\ (r_1, r_2) \\ (s_1, s_2) \end{bmatrix},$$

associated with a design in which two columns are missing, be defined to mean: treatments p_1 and p_2 are missing from row 1, treatments q_1 and q_2 are missing from row 2, etc. It is clear that the values of a_{i1}, a_{i2}, a_{i3}, that is, the members of the ith row of the a-matrix, depend only on the treatments missing from row i of the design. There are only ten different pairs of treatments which may be absent from any row, and therefore there are only ten different rows which may occur in the a-matrix. No advantage is gained here by reordering rows to bring the missing treatments in one of the columns into an assigned order since the number of different pairs is the same in any case. These ten pairs, together with the corresponding rows of the a-matrix, are listed below.

(7.1)

(1, 1)	-2	$+2$	$+2$
(1, 2)	-2	0	0
(1, 3)	0	$+2$	0
(1, 4)	0	0	$+2$
(2, 2)	-2	-2	-2
(2, 3)	0	0	-2
(2, 4)	0	-2	0
(3, 3)	$+2$	$+2$	-2
(3, 4)	$+2$	0	0
(4, 4)	$+2$	-2	$+2$

The setting-up of the a-matrix will consist simply in writing down the symbol associated with the design and substituting for each curved

bracket (p_1, p_2) etc., the corresponding row taken from this list. The use of this device is illustrated in section 9.

Similar lists can be made up for obtaining the a-matrix when more than two columns are missing, but the lists are long and presumably would be used only rarely.

8. Case (5). Suppose now that one column and another single observation are missing. This case can be treated by a combination of the methods used in sections 4 and 6. After rows have been reordered to bring treatment 1 in the missing column into row 1, etc., let the single missing observation occur in row l and column m. Substitute for this missing observation the symbol y_{lmn} and fit to the "completed" set of observations a regression equation $Y_{ijk} = m + r_i + c_j + t_k$. The values of the fitted constants are easily determined and, when substituted in the regression equation, yield the relation

$$Y_{ijk} = C_j/4 + [q(R_i + T_k) + R_k + T_i]/(q^2 - 1) - G/2(q - 1).$$

Now, according to the argument used in section 4, the number to be substituted for the missing observation is determined by equating y_{lmn} to Y_{lmn}. Thus the formula for the missing value is obtained by solving for y_{lmn} the equation

$$y_{lmn} = C_m/4 + [q(R_l + T_n) + R_n + T_l]/(q^2 - 1) - G/2(q - 1).$$

Two cases arise. If $n = l$, all terms on the right side of this equation involve y_{lmn}; if $n \neq l$, the terms R_n and T_l do not contain y_{lmn}. The solutions in the two cases are

$$(8.1) \qquad y_{lml} = [(q - 1)C_m' + 4(R_l' + T_l') - 2G']/(3q - 6),$$

$$(8.11) \qquad c_{11} = c_{22} = c_{33} = \frac{1}{4(q^2 - 1)}\left(q + \frac{q + 1}{3q - 6}\right),$$

and

$$(8.2) \qquad \begin{aligned} y_{lmn} = {} & [(q^2 - 1)C_m' + 4q(R_l' + T_n') \\ & + 4(R_n + T_l) - 2(q + 1)G']/(3q^2 - 6q - 1), \end{aligned}$$

$$(8.21) \qquad c_{11} = c_{22} = c_{33} = \frac{q}{4(q^2 - 1)}\left(1 + \frac{1}{3q - 6}\right), \text{ approximately.}$$

The prime indicates, as in section 4, that the total is incomplete. The derivations of formulae (8.11) and (8.21) are discussed later.

The argument given in section 4 shows that when this number is substituted for the missing observation and the analysis carried out

using the formulae of section 6, the values of the fitted constants and of the residual sum of squares are identically the same as would be obtained by fitting the regression equation to the original observations alone. However, the formulae given in section 6 for the variances and covariances do not hold here. The following considerations yield the proper values without much computation.

Suppose that the regression equation (5.1) is fitted to the original observations. The left sides of the normal equations will be considerably different from those of the equations in section 6, but the right sides, if changed at all, will be altered only by the addition or subtraction of the number y_{lmn}. Again using a prime to indicate that y_{lmn} is omitted, the right sides of these normal equations may be written G', $C^{*\prime}$, $R^{*\prime}$, etc., where $G=G'+y_{lmn}$, $R_1^*=R_1^{*\prime}\pm y_{lmn}$, the plus sign to be used if $l=1$ or 2 and the minus sign if $l=3$ or 4, with similar relations between the other primed and unprimed symbols. Now, to find the variance of one of the $t^{*\prime}$s, say t_2^*, we may replace $T_2^{*\prime}$ by unity and all other $T^{*\prime\prime}$s, $R^{*\prime\prime}$s etc. by zeros in the formula for t_2^* given by the solution of the normal equations. This substitution gives the value of c_{22}. Similarly, replacing $T_1^{*\prime}$ by unity and all other $T^{*\prime\prime}$s etc. by zeros, we get the value of c_{12}.

The formulae for the $t^{*\prime}$s are given in section 5 expressed in terms of the unprimed symbols. These formulae can be written in terms of the primed symbols (remembering that y_{lmn} is itself a function of the primed symbols), and thus, making the substitutions indicated above, are obtained the values of the c's.

The formulae for the c's do not fall into any simple pattern, but there are two main types corresponding to the two missing-value formulae.

When formula (8.1) is used, that is, when a treatment is missing twice in the same row,

$$c_{11} = c_{22} = c_{33} = \frac{1}{4(q^2-1)}\left(q + \frac{q+1}{3q-9}\right),$$

and c_{12}, c_{23}, c_{31} are equal to $\pm 1/12(q-1)(q-3)$. The arrangement of plus and minus signs depends on the whole configuration of missing treatments, but in all cases two of the covariances are positive and one is negative.

When formula (8.2) is used, the values of c_{11}, c_{22}, c_{33} are given by

$$\frac{1}{4(q^2-1)}\left[q + \frac{(q\pm 1)^2}{3q^2-6q-1}\right],$$

the plus sign applying to only one of the three, and the set of quantities c_{12}, c_{23}, c_{31} may take various combinations of the values

TABLE I

THE NUMBER OF THE TREATMENT IS RECORDED UNDER EACH OBSERVATION

Rabbits	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Totals columns 1–16	Totals columns 1–15	Totals columns 1–14	Totals columns 1–13
Days																				
1	59	56	45	62	42	49	83	56	47	90	79	50	45	52	57	64	936	872	815	763
	3	2	1	4	1	3	2	4	2	1	3	4	4	3	1	2				
2	56	58	41	49	39	61	81	54	46	74	63	69	61	31	30	83	886	803	773	742
	1	4	3	2	2	1	4	3	4	3	2	1	1	2	4	3				
3	41	73	30	63	44	38	101	65	62	61	58	66	45	35	57	74	913	839	782	747
	2	3	4	1	4	2	3	1	1	2	4	3	3	4	2	1				
4	54	69	28	84	61	43	96	58	76	63	87	59	71	81	50	67	1047	980	930	849
	4	1	2	3	3	4	1	2	3	4	1	2	2	1	3	4				
Totals	210	246	144	258	186	191	361	233	231	288	287	244	222	199	194	288	3782	3494	3300	3101

Treatments	Totals columns 1–16	Totals columns 1–15	Totals columns 1–14	Totals columns 1–13
1	1078	1004	947	866
2	845	791	724	693
3	1047	964	914	862
4	812	745	715	680

The data in Table I were provided through the courtesy of Dr. D. M. Young and the Connaught Laboratories, University of Toronto.

$\pm 1/4(3q^2-6q-1)$, $\pm(q-1)/4(q+1)(3q^2-6q-1)$. Thus it appears that the variances and covariances require a rather complicated set of formulae. The complexity may be avoided, however, by introducing suitable approximations. For any values of q that are likely to occur in practice $(q\pm 1)^2/(3q^2-6q-1)$ may be replaced by $q/(3q-6)$ without sensible loss of accuracy, which gives the approximate formula (8.21). Likewise, in such cases, the covariances are numerically small and may be ignored.

An argument of the same kind, applied to the case discussed in section 4, shows that the effects of a single missing observation are to increase the variances of B, C, D from σ^2 to $\sigma^2(12r-5)/(12r-6)$ and to change the covariances from zero to $\pm\sigma^2/(12r-6)$. These changes are seen to be so small numerically as to permit the use of the formulae appropriate to the intact design, ignoring the slight inaccuracies resulting from the missing observation.

When one column and several single observations are missing, formulae (8.1) and (8.2) may be used iteratively to supply values for the single observations. An analysis performed on this "completed" set of observations, using formulae (6.1) and (6.2), yields the correct values of the residual sum of squares and of t_1^*, t_2^*, t_3^*. The variances and covariances of the t^*'s are affected by the missing observations, but the error committed in using formulae (8.11) and (8.21) in such cases should be small, unless the number of missing values is large.

9. Examples. Case (1). Suppose that the observation in row 1 and column 16 of Table I is missing. Then $R_1'=872$, $C_{16}'=224$, $T_2'=781$, $G'=3718$. These numbers, substituted in formula (4.1), give the number to be used for the missing observation.

$$y_{1(16)2} = [16(224) + 4(872 + 781) - 2(3718)]/42 = 66,$$

to the nearest integer.

The analysis of the completed table, carried out as in section 1, gives the following values for the sums of squares and the components B, C, D.

	d.f.	s.s.	m.s.
rows	3	935	
columns	15	10,380	
treatments	3	3,461	
residual	41	2,587	63.11
total	62	17,363	

$$D = 8.25, \qquad B = -58.25, \qquad C = 0.50$$

$$M = (.30103)(8.25)/(-58.25) = -.0426$$

$$s_M = (63.11)(.30103)\sqrt{(58.25)^2 + (8.25)^2}/(58.25)^2 = .0329.$$

Here the value of M is correct, but the formula and value of s_M are inaccurate, owing to the correlation of B and D and the altered values of their variances. Correcting for both of these sources of error changes the value of s_M to .0335, a negligible correction, particularly in view of the fact that the formula for s_M is itself only approximate. Similarly, corrections in the formulae for testing the significance of B and C are not important and may be ignored.

Case (2). Suppose that columns 14, 15 and 16 are missing. The first step in setting up the normal equations is the computation of the a-matrix. The value of a_{23}, for example, is the sum of the values taken by w_3 for the 13 observations in row 2. Now w_3 takes the value 1 when the observation is made on treatments 2 or 3, and takes the value -1 when the observation is made on treatments 1 or 4. The treatments missing from row 2 are treatments 2, 4, 3. Hence $a_{23}+1-1+1=0$, $a_{23}=-1$, since each w sums to zero over each row when the design is intact. In this way (or others) the a-matrix may be set up in a few minutes. It is given below, multiplied on the left by the u-matrix to form the A-matrix.

$$\begin{pmatrix} 1 & 1 & -1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} -1 & 1 & -1 \\ 1 & -1 & -1 \\ -1 & -1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & -4 \\ 4 & 0 & 0 \\ 0 & -4 & 0 \end{pmatrix}.$$

The right sides of the normal equations (the R^*'s and T^*'s) are computed from the row and treatment totals.

$R_1 = 763$	$T_1 = 866$	$R_1^* = -91$	$T_1^* = +17$
$R_2 = 742$	$T_2 = 693$	$R_2^* = +81$	$T_2^* = -355$
$R_3 = 747$	$T_3 = 802$	$R_3^* = -123$	$T_3^* = +0$
$R_4 = 849$	$T_4 = 680$		

The calculation of the R^*'s and T^*'s may be checked by means of the identities

$$R_1^* + R_2^* + R_3^* + G = 4R_2, \qquad T_1^* + T_2^* + T_3^* + G = 4T_3$$

When the normal equations are set up, it is found that the six equations break up into three pairs. The solutions are

$$672t_1^* = 13T_1^* - R_2^* \quad 672t_2^* = 13T_2^* + R_3^* \quad 672t_3^* = 13T_3^* + R_1^*$$

$$672r_2^* = 13R_2^* - T_1^* \quad 672r_3^* = 13R_3^* + T_2^* \quad 672r_1^* = 13R_1^* + T_3^*$$

Substitution of the numerical values of the R^*'s and T^*'s yields

$$t_1^* = 0.2083 \qquad r_1^* = -1.7470$$
$$t_2^* = -7.0506 \qquad r_2^* = 1.5417$$
$$t_3^* = 0.0387 \qquad r_3^* = -2.9077$$

Substitution of unity for T_1^* and zeros for all other T^*'s and R^*'s in the formula for t_1^* gives $c_{11}=13/672$. Similarly, replacing T_2^* by unity and all other T^*'s and R^*'s by zeros in the same formula gives the value of c_{12}. Evidently,

$$c_{11} = c_{22} = c_{33} = 13/672, \qquad c_{12} = c_{23} = c_{31} = 0.$$

The reduction in the sum of squares due to rows and treatments is

$$\sum r_\alpha^* R_\alpha^* + \sum t_\gamma^* T_\gamma^* = 3148.$$

The other numbers needed to complete the analysis of variance table are: the "total" sum of squares, $(59)^2 + \cdots + (71)^2 - (3101)^2/52 = 198533 - 184926.94 = 13606$; the "columns" sum of squares, $[(210)^2 + \cdots + (222)^2]/4 - (3101)^2/52 = 8847$.

	d.f	s.s	m.s
columns	12	8,847	
rows and treatments	6	3,148	
residual	33	1,611	48.80
total	51	13,606	

t_2^* is highly significant, since

$$F = \frac{(t_2)^{*2}}{s^2 c_{22}} = 52.65.$$

Likewise t_3^* is not significant.

$$M = -.0889 \qquad \text{(Formula (5.3))}$$
$$s_M = .0411 \qquad \text{(Formula (5.4))}$$

Case (3). Suppose that column 16 is missing. Reordering rows to bring treatment 1 of column 16 into row 1, etc., in order to apply the formulae of section 6, the row and treatment totals and the numbers calculated from them are:

$R_1=839$	$T_1=1004$	$R_1^*=-72$	$T_1^*=+76$	$R_1^*+T_1^*=+4$
$R_2=872$	$T_2=781$	$R_2^*=+210$	$T_2^*=-442$	$R_2^*+T_2^*=-232$
$R_3=803$	$T_3=964$	$R_3^*=-144$	$T_3^*=-4$	$R_3^*+T_3^*=-148$
$R_4=980$	$T_4=745$			

The numbers t_1^*, t_2^*, t_3^*, given by formula (6.1) are 1053, -0420, -204, divided by 896. The reduction in sum of squares due to rows and treatments is

$$[(-72)^2 + \cdots + (-4)^2]/64 + [(4)^2 + \cdots + (-148)^2]/896 = 4321.67. \qquad \text{(Formula (6.2))}$$

The other numbers needed for the analysis of variance table are calculated as in the preceding example.

	d.f	s.s	m.s
columns	14	9,616	
rows and treatments	6	4,322	
residual	39	2,477	63.52
total	59	16,415	

The tests of significance and the computation of M and its standard error proceed as in the preceding examples, using formula (6.11) to obtain the value $c_{11} = c_{22} = c_{33} = 0.01674$.

Case (4). Suppose that columns 15 and 16 are missing. The symbol describing the distribution of missing treatments and the a-matrix derived from it by means of list (7.1) are

$$\begin{Bmatrix} (1,\ 2) \\ (4,\ 3) \\ (2,\ 1) \\ (3,\ 4) \end{Bmatrix} \quad \text{and} \quad \begin{Bmatrix} -2 & 0 & 0 \\ +2 & 0 & 0 \\ -2 & 0 & 0 \\ +2 & 0 & 0 \end{Bmatrix}. \quad \text{The } A\text{-matrix is } \begin{Bmatrix} 0 & 0 & 0 \\ 8 & 0 & 0 \\ 0 & 0 & 0 \end{Bmatrix}.$$

The rest of the calculation follows exactly as in case (2).

Case (5). Suppose that column 16 and the observation in row 1, column 15 are missing. Again the rows are reordered as in case (3) and the missing observation now occurs in row 2. The following numbers are needed to apply formula (8.2) to supply a number for the missing observation.

$$C'_{15} = 137, \ R_2' = 815, \ T_1' = 947, \ R_1 = 839, \ T_2 = 781, \ G' = 3437.$$

Substituting these numbers in formula (8.2),

$$y_{2(15)1} = [(224)(137) + 60(815 + 947) + 4(839 + 781) - (32)(3437)]/(584) = 56.34$$

or 56 to the nearest integer.

Substituting 56 for the missing observation, the analysis proceeds exactly as in case (3), with 38 instead of 39 degrees of freedom for the residual sum of squares and using formula (8.21) instead of (6.11) to

obtain the values of the c's. The value so calculated (0.01717) differs only slightly from that given by formula (6.11) (0.01674), but the more accurate formula (8.21), appropriate to this case, requires only simple calculations and presumably should be used.

Conclusion. The primary purpose of this paper is to bring together a number of methods of dealing with the analysis of balanced designs when some observations are missing. The discussion is focussed on a specific design, but the methods and devices can be used, with suitable modifications, in a much wider range of problems.

Another purpose may also be served. Apparently some suspicion attaches to the "estimation of missing values" as being an attempt to extract something from nothing. Perhaps the phrase itself, although warranted, is misleading. In any event, it is hoped that the foregoing discussion emphasizes sufficiently the fact that the basis and method of analysis are the same in all cases, and that the cases singled out here for discussion are special only in the sense that, in them, some simplification of the arithmetic is possible.

REFERENCES

[1] Yates, F., "Incomplete Latin Squares," *Jour. Agric. Sci.*, Vol. XXVI, 1936.
[2] Yates, F. and R. W. Hale, "The Analysis of Latin Squares When Two or More Rows, Columns, or Treatments Are Missing," *Supplement to Jour. Roy. Statistical Soc.*, Vol. VI, 1939.
[3] Yates, F., "The Analysis of Replicated Experiments When the Field Results Are Incomplete," *Empire Jour. Experimental Agric.*, Vol. I, 1933.
[4] Bliss, C. I. and H. P. Marks, "The Estimation of Drug Potency from a Graded Response (II)," *Quarterly Journal of Pharmacy and Pharmacology*, Vol. XII, No. 2, 1939.
[5] Goulden, C. H., *Methods of Statistical Analysis*. New York, John Wiley and Sons, 1939.

DR. VICTOR SELDEN CLARK*

WITH the death of Dr. Victor Selden Clark in the 77th year of his age, the country has lost one of its outstanding economists. He was not concerned either with economic theory or with statistical methods. He was primarily an economic analyst, making use of all available-sources of information concerning the several subjects which he investigated and interpreting them in masterly fashion.

Through much of his career, Dr. Clark devoted his attention mainly to economic conditions and trends in those areas of the world concerning which adequate information was not readily available—outlying possessions of the United States, Latin American and Far Eastern countries, and the "down under" lands of Australasia. Most of his time from 1900 to 1913 was spent on studies of these areas. The emphasis was on labor problems but the whole economic picture of each area was sketched. In no case did he rely mainly on published information, official or unofficial; he visited each area and obtained most of his knowledge of it from personal observation and interviews with all classes of people. He may well be characterized both as a "globe trotter" and as a past master of the reporter's art. The areas covered by his researches included Puerto Rico and Hawaii (in both of which he for some time held high-ranking administrative positions), Philippine Islands, Cuba, Mexico, Java, New Zealand, Australia, and Canada; incidentally he visited Russia and Spain. The results of this work were for the most part published by the Bureau of Labor Statistics. Each of the monographs which Dr. Clark prepared for that Bureau contains from 100 to 150 closely-packed pages and was, at that time, the principal source of information on the given area. They included critical discussion of problems in addition to factual material.

Toward the end of his career, Dr. Clark reverted again to globe-trotting, not only revisiting practically all those areas where he had previously made investigations, but also travelling extensively in other countries of the Far East and Latin America. For the most part, the results of these later travels and researches were not published. However, the investigation concerning Puerto Rico, which he and colleagues conducted under the auspices of the Brookings Institution during the years 1928–30, constituted the basis for a large volume *Puerto Rico and Its Problems*, which is the most valuable one source of informa-

* Born at Portageville, N. Y., June 12, 1868, son of Major Selden N. and Helen E. (Davis) Clark. Litt.B., U. of Minn. 1890; Ph.D. Columbia, 1900; died April 1, 1946, Washington, D.C.; unmarried.

tion ever compiled concerning that island and its difficult economic, social, and political problems.

In the interval between these two periods of research in outlying countries, Dr. Clark conducted a monumental investigation concerning the economic history of this country itself, under the auspices of the Carnegie Institution. The first volume, *History of Manufactures in the United States, 1607–1860*, was published in 1916. The work was then interrupted by the war and the second volume covering the period 1860–1914 was not issued until 1927. The two volumes together contain nearly 2,000 pages and were exhaustive in scope and illuminating in analysis. Professor Henry W. Farnan said in his introduction to the second volume:

> . . . With rare persistence and industry he has succeeded in completing the work which is now offered to scholars. Like the History of Manufactures to 1860, this study is based on original material. It is an economic history in the strict sense of the word. It does not deal with technology and mechanics; it does not give biographies of prominent manufacturers; it can not cover the details found in histories of specific industries. It does give an interpretation in broad outlines of the development, the organization and the economic interactions of manufacturing industry in our country during a truly remarkable period.

In preparing this history of American manufactures, Dr. Clark made use of an immense variety of material, ranging from official statistics of the Federal and State governments and of trade associations to items in periodicals and personal reminiscences. He traced the effects of tariffs on the development of American industry, the gradual spread of manufactures from the northeast to the central, southern and western sections of the country, the progress of the working classes, the development of great corporations and combinations, and all the other major aspects of the history of industry.

In *What Is Money*, a compact 88-page book published in 1934—a time when the country was passing through a maze of monetary discussion—Dr. Clark gave an elementary explanation of money and its relation to prices. His approach to the subject, as he said in his preface, was "by the path of history rather than by that of theoretical analysis." Attractively written, with frequent touches of humor, his little book might well continue to be a valuable guide for the "man in the street" or non-professional reader.

During the years 1920–28 Dr. Clark was editor of *The Living Age*, Boston, a periodical of long standing and wide circulation, which, however, was ultimately discontinued as other "readers' digests" became more popular. During his later years Dr. Clark maintained an

office in the Library of Congress, where he not only pursued his own investigations but served as a consultant to the Library regarding economic and governmental publications.

E. DANA DURAND
Washington, D. C.

CORRINGTON CALHOUN GILL

1898–1946

THE death of Corrington C. Gill in Tucson, Arizona, on July 13, 1946, closed a distinguished career as government economist, statistician and administrator. Gill played a significant part in shaping and executing governmental policy during the long depression and the war period.

Gill was born in Grand Rapids, Michigan, in 1898. His early life was spent in Michigan. Entering the Navy in 1917, he served throughout the war in the destroyer service on the French coast. He received his A.B. degree at the University of Wisconsin in 1923.

In 1923 he came to Washington as Manager and Correspondent of the Washington D. C. Press Service. After four years he turned to independent research and business consulting work. In 1931 he became economist and statistician for the Federal Employment Stabilization Board.

In the Spring of 1933 Gill was selected by Harry Hopkins to direct research, statistical and finance activities for the newly-created Federal Emergency Relief Administration. He occupied a key position in federal relief agencies for the following eight years, serving successively as Director of the Division of Research, Statistics and Finance of F.E.R.A.; Assistant Administrator of F.E.R.A.; Assistant Administrator of the Civil Works Administration; and Assistant Commissioner of the Work Projects Administration. The excellent administrative record of the emergency organizations during this period was in large measure the result of his efforts.

In 1941 Gill left W.P.A. to become Deputy Director of the Office of Civilian Defense in charge of Operations. The difficult job of organizing regional, state and local civilian defense operations throughout the nation was performed with his usual speed and effectiveness. In 1942–1943 he served for a year as Consultant to the War Department, making a complete survey of the Army Medical Corps and assisting in its reorganization.

In the Spring of 1943 Gill was chosen by the President to become Director of the Committee for Congested Production Areas, assisting localities with war-swollen populations to obtain urgently needed facilities and community services. When the C.C.P.A. was liquidated, he became Deputy Director General of the United Nations Relief and Rehabilitation Administration in charge of Administration and Finance. He held this position until late 1945 when illness necessitated his resignation.

He is survived by his widow, Julia Turnbull Gill. He is the author of numerous economic and statistical articles and of two books, *Wasted Manpower*, and *The Challenge of Unemployment*.

Those who knew Corrington Gill and worked with him shall miss him a great deal. An unusually competent administrator, he had a rare ability to bring out the best in those who worked under him. Numerous new developments and techniques in research and statistics were developed under his direction and with the assistance of his wise judgment and shrewd sense of timing. His ability to complete stupendous and urgent tasks competently and quickly is rare, in Washington as elsewhere.

HOWARD B. MYERS
Committee for Economic Development

BOOK REVIEWS

Edited by

Oscar Krisen Buros

Rutgers University

Statistical Abstract of the United States, 1944–45: Sixty-Sixth Number. Compiled by the Bureau of the Census. Washington 25: Government Printing Office, 1945. Pp. xiii, 1023. $1.75.

Review by Bruce D. Mudgett
Professor of Economics and Statistics
University of Minnesota

On first thought a review of the *Statistical Abstract of the United States* seems almost in a class with writing a literary account of the monthly telephone bill, and the present writer had that reaction upon being asked to write this note. But maybe on further consideration there is something to be said, if not too much, about a recurrent annual volume summarizing the vast quantities of statistical information that pour out each year from and about our national government and our national economy.

It should be said at the start that the present volume is the current result of a development or evolution of many years (66 in fact) in presenting a single volume of brief summary information that is in wide demand. Of the things that would seem to deserve notice in a technical comment, I do not conceive the factual content of the volume deserves a very important place—and that mainly because most people who will have occasion to use the *Statistical Abstract* are likely to have a good general idea of the sort of information that is available, information such as is collected and published by different subdivisions of government and by some non-governmental agencies. Illustrations: population, vital statistics, prices, foreign and domestic trade, agricultural, manufacturing and mining production and the like. There have been additions to the factual content as the years have passed and a cursory examination of the present volume indicates that some of the recent additions have come from the sample studies developed in the last few years. For example, section five on the labor force contains tables constructed from data obtained in the monthly report on the labor force—a very new and very important type of inquiry now being carried on in the Census Bureau.

Since the *Statistical Abstract* is a compendium of summary statistical information in a vast number of fields, the most important thing about it would seem to be how well has it done the job it is designed to do. This is a matter, largely, of the organization of the book and of the construction of the individual tables so that users can find what is there and can understand it when found. A new feature in this volume, and a commendable in-

novation is "the presentation of a general note at the beginning of each section in order to furnish a background for the statistics in the field—." Such notes have been prepared for fifteen of the thirty-four sections and further extensions are promised in future editions. This feature plus one other makes for easy location of available information on any topic. The other is an extensive and carefully constructed index. Furthermore the source reference for each table (of which there are over one thousand) is an improvement over what used to be, in that the source reference gives the publication in which the original data appeared, as well as the agency which published it. For example, table 80 of this volume giving death rates per 1,000 population by sex and age-groups 1900–1943 lists the source as "Dept. of Commerce, Bureau of the Census; Vital Statistics Rates in the United States 1900–1940; basic figures 1940–1943, annual report, Vital Statistics of the United States, Part II." In contrast the source reference for the corresponding table of the 1931 Abstract (there it is table 76) merely lists "Bureau of the Census, Dept. of Commerce." In the front of the volume is a list of all tables classified by source; thus all tables obtained from the Bureau of Agricultural Economics are listed under that title in this classification. This year's practice on source reference is a great improvement for anyone who wishes to seek more complete data in the original source.

The feature of most importance in the construction of individual tables is the appropriate arrangement of classifications in stub and caption and the adjustment of spacings, lines, differences in type, etc., in such a way as to bring out the information contained in the figures. We have all seen examples of badly constructed tables where the effort was almost more than its worth to discover what the figures meant. In this respect the *Statistical Abstract* has for many years done a reasonably good job, which could in at least one respect be improved; but I suspect the improvement would run counter to long-established habits in the Government printing office and therefore, would have a great inertia of habit to overcome before it could be introduced. I refer particularly to the use of different weights of line and sometimes differences in type in tables to help to show coordinate and subordinate relationships among the figures. There is evidence of some tendency in this direction, but it could go much further, to the great improvement of the abstract. Consider for example, table 289 on Income Tax Returns to States and Territories for 1939, 1940, 1941. In the stub classification, U. S. totals and section designations are in heavy faced type, whereas individual states (subdivisions of the sections) are in lighter type, equally for the figures referring to these several categories. But it should be noted that in some tables the heavy faced type is scarcely distinguishable from the lighter faced variety. This, of course, is a printer's problem, but it does deserve more attention than it has received in our government printing office. Statistical volumes in many European countries have for years been vastly better than ours on this score. Again illustrating from table 289, these tables make much less use of different weights of line than is desirable,

and again the proof lies in the better job done in European printing offices. Table 289 would be greatly improved if, for example, the lines between 1939 and 1940 were much heavier than the two lines within the 1939 category; and similarly for other years.

The above criticism that the *Statistical Abstract* does not make enough use of differences in type and different weight of lines is, I think, deserved, and here is one chance for improvement on the technical side of presentation. Despite this objection, the government has placed within some thousand pages a vast body of summary statistical data and has organized it in such a way as to make it readily available to its potential users—a truly monumental achievement.

Statistical Analysis for Students in Psychology and Education. *Allen L. Edwards* (Associate Professor of Psychology, University of Washington). New York 16: Rinehart & Co., Inc. (232 Madison Ave.), 1946. Pp. xviii, 360. $3.50.

REVIEW BY DAVID A. GRANT
Associate Professor of Psychology
University of Wisconsin

EDWARDS' *Statistical Analysis* is not just another elementary statistics book. Unusual features and clean breaks with tradition are to be found in both structure and content, and several of the unique features are to be applauded. For example, one-third of the book is given over to the statistics of small samples, and *two* chapters are devoted to analysis of variance. The reviewer agrees with Edwards that " . . . whether the traditional attitude approves or not, more and more research as published in psychological and educational journals is being evaluated by small sample techniques." Other good points about *Statistical Analysis* include clearly written elementary chapters on "Probability and Frequency Distributions" and "Research and Experimentation," modern sampling treatment for the correlation coefficient, the use of the Charlier checks in computation, and a chapter on elementary algebra. This last-mentioned chapter is perhaps a bit too elementary omitting logarithms and exponents. Some teachers of statistics may have mixed feelings about teaching statistics to students whose mathematical backgrounds are so inadequate as to permit them to profit from such a simple review of algebra.

One can scarcely approve of one of the unusual features of the book, the deliberate omission of any treatment of the reliability of test scores. To begin with, this material has unique value for students in psychology and education. Its absence detracts materially from the usefulness of *Statistical Analysis* as a general elementary textbook in statistics for psychologists and educators. Furthermore, in spite of Edwards' comment that it does not " . . . fall within the general orientation of the rest of the book . . . , " fallibility is a property of *all* actual mensuration, including reflex magnitudes,

discrimination limens, reaction times, and other measures from "pure" psychology. To omit the topic of reliability even from an experimentally oriented statistics book such as *Statistical Analysis* would seem to encourage superficial interpretation of the nature of measurement in general.

In the reviewer's opinion, *Statistical Analysis* suffers from a number of pedagogical weaknesses. These include: (1) an undesirable multiplicity of definition and computational formulae which seems bound to confuse the elementary student; (2) instances of unnecessary notational complications, such as the use of v for σ^2, and z for the critical ratio or significance ratio which requires the following footnote (p. 180): "Fisher has introduced a transformation of r into another statistic which is known as z (not to be confused with the z mentioned earlier) . . . " ; and (3) introduction of Peters and Van Voorhis' ϵ^2 as an alternative to F in analysis of variance, accompanied by the statement, "You may wonder at this point whether F or ϵ^2 should be used as a test of significance in a given problem. There is no definite answer to this question. . . . " (p. 236). It seems to the reviewer that an elementary textbook should avoid such sources of confusion.

The reviewer's chief criticism centers around the general emphasis of the book. The reader, intrigued by the list of chapter topics, and led on by a promise in the preface that emphasis is to be upon use and function rather than calculation, is doomed to a real disappointment as he works his way through calculation after calculation, coding technique after coding technique. On the topic of correlation, for example, 12 pages are expended on the development and use of definition and computational formulae, but only two and one-half pages are given to interpretation of r, and most of this is rather superficial. In this respect, it would seem that Edwards is caught up in precisely the very "vicious circle" he promises, in the preface, to break.

With respect to statistical errors, Edwards' book is very much like the first editions of most elementary statistics texts which reach the market. The slips are disconcerting, but minor. They include: (1) the statement that computation of the probable deviation requires computation of the S.D. (p. 60); (2) an inaccurate technique used to obtain fiducial limits for a proportion (p. 169); (3) a remark that grouping will cause no serious error in the S.D. (p. 70); (4) a formula for t which holds only when $n_1=n_2$ (p. 182); (5) an erroneous assertion about σ_r (p. 185); and (6) an interpretation which accepts the null hypothesis when χ^2 does not exceed the 5 per cent point (p. 247).

Statistical Analysis is better than many current statistics books. But, because of the points mentioned above, the reviewer does not think that it will prove to be an excellent teaching text. He does not rate it in the same class with the elementary texts of Lindquist and Helen Walker. Edwards' book is, in many respects, a more forward-looking text, but it lacks the teaching finesse and the depth of interpretation of its best available competitors.

Income from Independent Professional Practice. *Milton Friedman* (Associate Professor of Economics, University of Chicago) and *Simon Kuznets* (Professor of

Statistics and Economics, University of Pennsylvania). Publications of the National Bureau of Economic Research, No. 45. New York 23: National Bureau of Economic Research, Inc. (1819 Broadway), 1945. Pp. xxxiii, 594. $4.50. *Two reviews follow:*

REVIEW BY R. L. ANDERSON
Associate Professor of Experimental Statistics and Agricultural Economics, Institute of Statistics University of North Carolina, Raleigh

THIS book summarizes the results of several years' study of the net income from independent practice by physicians, dentists, lawyers, certified public accountants and consulting engineers. The authors have made use of many modern statistical techniques such as the χ^2-test, the analysis of variance and multiple regression, which in the past have been applied too infrequently to the analysis of economic data. Dr. Friedman's techniques for the analysis of ranked data have also been utilized to good advantage in several places.

Because of the wide usage of modern statistical techniques, this book should receive careful scrutiny by economic statisticians. The authors are careful to point out the basic assumptions behind such tools as the analysis of variance, assumptions which are too often not fulfilled with economic data. Such a treatment should help immeasurably to promote the idea of finding means of altering the old statistical tools or of devising new ones in order that economic data can be analyzed more objectively than in the past. While some criticisms will be made below of the statistical methods employed, these criticisms in the main are minor in character and the general quality of the statistical methodology is well worth the attention of economists.

The first two chapters describe the professions studied and the methods used in obtaining the samples and in adjusting for known biases in the sample results. Unfortunately, the authors did not have available some of the modern techniques of conducting mailed questionnaires, which were used exclusively in this study. It was concluded that the nonresponse problem was not important in this survey, because no correlation was found between income and response. However, the geographical distributions of the sample responses did not correspond with those of the population. It is the reviewer's opinion that such an extensive study should have been based on a more solid sampling foundation with some provision made for sampling the nonresponders by direct interview. Also some provision should have been made to check on the accuracy of actual replies. For example, did deductions from gross income include only business and not personal expenses?

The authors present a detailed analysis of the adjustments necessitated by such things as the exclusion of all but members of the American Dental Association in the dentist's sample and the apparent size of community bias in the samples for doctors and lawyers. The sample and population frequencies were compared by use of the χ^2-test for goodness-of-fit and of χ_r^2, the latter being Friedman's test for rank correlations. The attempts to justify

the results on the basis of previous state surveys seemed inconsistent with the previous statement that the sample results were not good for separate geographical units.

Chapters 3 and 4 contain the pertinent conclusions of the study, comparisons between professional and other income and comparisons of income among the five professions. Average professional income was about three times as large as that for people not engaged in one of these professions. This result was largely explained by the more expensive and longer training required of professional men, the larger communities (with larger average income for any work) which attracted professional men, the higher ability required for professional work, and the noncompetitive nature of the professions. My opinion is that too little importance was placed on the factor of ability. However, the restriction of entry and the noncompetitive nature of the professions cannot be stressed too much. These are factors for which an adequate statistical analysis is not available; hence, the possibility for personal bias in the evaluation of causes is quite important. One also wonders what the inclusion of educators would have done to the over-all professional picture as herein presented.

For all of the income data, it was found that the average incomes and their standard deviations were highly correlated, indicating that it might have been better to assume logarithmic relationships in the analyses. Lorenz curves were used to study relative variability among professional incomes and between professional and nonprofessional incomes. The five professions ranked as follows as regards average income (high to low): engineers, accountants, lawyers, doctors and dentists. There was little difference between the average income for lawyers and doctors.

A detailed analysis was made of the difference between dentists' and doctors' incomes. This reviewer tends to agree with a comment made by C. R. Noyes (pp. 405–10) to the effect that the assumptions made about the differences between these two professions were of such an arbitrary nature as to render highly untenable the statement that doctors restrict entry into their profession more than dentists. A multiple regression was run for the doctors' average income in each state using the percentage of doctors in the population and the per capita income as independent variables; similarly, for dentists. This kind of analysis of economic data should be used much more. In this case the authors were able to obtain the elasticity of demand for services for a given change in the percentage of doctors and dentists in the community, holding per capita income constant.

Chapters 5 and 6 present a detailed analysis of several causes of income differentials. The analysis of variance with the method of expected subclass numbers for disproportionate frequencies might have been profitably used to test for differences among regions and among communities of different sizes. Instead the authors utilized the χ_r^2 method mentioned above. Data were too scanty to discuss the important topic of influence of training and inherent ability on professional income. The importance of such intangibles as family connections and political influence are not stressed enough.

Chapter 7 discusses the stability of relative income status for the same individual in different years. A new treatment of the problem is presented here whereby income is said to consist of permanent, transitory and quasi-permanent components. If this analysis can be coupled with some estimate of the variability of the separate components, it might prove of value in locating the cause of the extreme variability in professional incomes. Linear regression analysis was used to compare incomes in successive years. This reviewer believes that any recommendations based on annual trends are extremely hazardous because of the correlations due to general business conditions. It is suggested that a multiple regression analysis utilizing some general business index as another independent variable is needed. The authors use the analysis of variance to test for nonlinearity. They are quite perturbed about the normality requirement for the analysis of variance, but the correlation of mean and variance is a much more important disturbing factor.

Chapter 8 presents some generalizations on the temporal changes in income for each profession and for business as a whole. The most important conclusion here is that there is a decided positive correlation between changes in professional and in general business incomes.

REVIEW BY ZENON SZATROWSKI
Instructor in Economics, Northwestern University

THIS study deals with the incomes from independent practice in five professions, medicine, dentistry, law, certified public accountancy, and consulting engineering; however, intensive analysis is applied to the more adequate data of medicine and dentistry. The data, obtained from questionnaires sent to sample groups, covers the period from 1929 to 1936. Presentation of these statistics in the volume, though a significant contribution, is only a small part of the contents. Analyzing differences in income and measuring the influence of factors giving rise to these differences appears to to be the principle objective. To quote the authors (p. 63): "But it is not enough merely to name factors making for variability in income. To be useful the catalog must be quantitative as well as qualitative; the importance of the different factors and the direction and magnitude of their influence must be measured."

Pursuing this objective, the authors use statistical techniques which are relatively simple but adequate for the problem. Pertinent characteristics of the distributions of incomes are obtained and used to describe differences in professions and variation from year to year. Arithmetic averages, medians, and quartiles are employed as estimates of general level of income; the interquartile range, standard deviation and Lorenz curves serve to measure variability in incomes. These statistical measures are used as a basis for comparing incomes from independent practice with incomes from salaried employment in the same professions, with incomes in other occupations and, of course, differences between the professions. Correlation techniques are used

to measure the stability of incomes from year to year and the relationships between income and its determinants, such as size of community, years of practice, etc. Index numbers are derived in order to describe temporal changes in the level and variability of incomes for the different professions. The reliability of the data itself is discussed at length. The objective tests of reliability which are used are based on a comparison of sample data with samples of other studies and the universe, with application of the χ^2 test.

The approach followed by the authors is to give a relatively complete discussion of the problem in connection with each step in their statistical analysis. This is desirable because it enables an individual with little statistical training to get a good deal of information about the subject. In addition, in this way, the authors present the reasoning which justifies their statistical procedure. For example, a priori they derive the determinants of income by considering the factors which influence the demand, the supply and the nature of the competition in connection with the different professions.

In their discussion the authors are careful to point out the limitations and shortcomings of their data and the statistical techniques they employ. The problem, from their standpoint, is to extract as much information as possible from the relatively inadequate data which is available. They do this competently. The value of the book lies not merely in the information obtained from the data which was available. The procedure which is developed can be used profitably as a pattern for both a further study of more complete data on the same professions, and also in connection with studies of other income groups.

Mathematics of Investment, Third Edition. *William L. Hart* (Professor of Mathematics, University of Minnesota). Bound with *Tables for Mathematics of Investment*, Third Edition. Boston, Mass.: D. C. Heath & Co., 1946. Pp. vii, 304; ii, 120. $3.00.

REVIEW BY LLOYD A. KNOWLER
Chairman, Department of Mathematics
State University of Iowa

THE text consists of the three parts: Part I, Annuities Certain; Part II, Life Insurance; and Part III, Auxiliary Topics. Essentially a fourth part is provided by a rather extensive set of tables for annuities certain and a few functions of the American Experience Table of Mortality with interest rate of $3\frac{1}{2}\%$.

The main features by which this edition differs from the preceding one are outlined in the preface somewhat as follows:

> An alteration has been made in the presentation . . . with the aim of giving the student maximum acquaintance with . . . applications before he meets the more unusual situations. . . . The exercises have been freshened. . . . The format of the text and tables has been changed. . . .

In carrying out "The primary aim to adapt the material, particularly that dealing with annuities certain, to the needs and the ability of the typical

student in a college of business administration" the author has attempted to keep as the "minimum prerequisite . . . a substantial second course in high school algebra" and as a "maximum prerequisite (with certain exceptions) a standard first course in college algebra."

Part I dealing with annuities certain covers pretty much the material in standard texts of mathematics of investment or finance. As with the preceding edition, it is rather well written. A nice collection of 106 review problems is given at the end of this part.

Part II dealing with life insurance consists of three chapters, namely: Life Annuities, Life Insurance, and Policy Reserves. On page 204 we find that "Table XIV (The American Experience Table of Mortality) was formed from the accumulated experience of many American life insurance companies." The accuracy of this statement is doubted in that it has been previously reported that the mortality statistics were deduced from experience of the Mutual Life Insurance Company of New York, but that the figures were inadequate at the older ages and so they were arbitrarily adjusted somewhat. Also, in the preface, we observe that "the notation and emphasis in Part II have been oriented thoroughly with respect to existing actuarial customs." Some sources might question the use of the American Experience Table at $3\frac{1}{2}\%$ interest as conforming to this custom. The use of this table at the interest rate indicated has no effect, however, upon the underlying principles. This part is meant, apparently, to be no more than a mere introduction to the elementary principles of the mathematics underlying life insurance.

Among the auxiliary topics in Part III is a discussion of computation and logarithms; progressions; and an appendix consisting of computation of interest by use of binomial expansion, force of interest, comparison of compound and simple interest for fractional interest period, certain interpolations, reference to depreciation charges, and abridged multiplication—a valuable aid which in the opinion of the writer should be used in those classes where appropriate tables or computing machines are not available.

Although several good textbooks exist in this area, it is worthwhile to consider this one for use in an introductory course. It will, no doubt, be adopted by a number of institutions.

Theory of Functions: Part 1, Elements of the General Theory of Analytic Functions. *Konrad Knopp* (Professor of Mathematics, University of Tübingen, Germany). Translated from the fifth German edition by *Frederick Bagemihl* (Instructor in Mathematics, University of Rochester). New York 19: Dover Publications (1780 Broadway), 1945. Pp. vii, 146. $1.25.

Review by Edmund Churchill
Instructor in Mathematics, Rutgers University

The dependence of modern statistical theory on advanced mathematical tools makes it increasingly desirable that the statistician have a background in mathematics that goes well beyond the traditional courses in

calculus and algebra. Some knowledge of the theory of functions of both real and complex variables, for example, is certainly essential to a complete understanding of such books as those by Kendall and Wilks on statistical theory and of most of the literature on distribution theory. The statistician not already acquainted with the elements of complex function theory will find this little book by Knopp an excellent introduction to the subject. It is a remarkably readable book, concise, clear, and rigorous. A number of brief illustrations and alternate forms of definitions help to clarify important points.

While this book was not written especially for the statistician, the material which it contains is so basic as to be valuable in its entirety to the statistician. The subjects treated consist of numbers and points, the general concept of a function of complex variable, integrals of complex functions, Cauchy's integral theorem and formulas, Taylor and Laurent series, analytic continuation, singularities, and the residue theorem. Perhaps half of these topics are involved, for example, in calculating the moments of the normal distribution by means of its characteristic function. The residue theorem often offers the simplest method of calculating integrals that arise in statistical problems. Other problems of statistical importance require a good deal more function theory than could be included in a book of this size, but a knowledge of what is in Knopp's book should enable one to comprehend most of the advanced theory in the more complete treatises when it is needed. What, in turn, is needed for an understanding of this book is roughly the equivalent of three semesters of calculus, including a good grasp of the concepts of convergence and continuity. As a book for self study, this book suffers from the small number of exercises included. A person using the book for this purpose would do well to obtain the first volume of Knopp's *Aufgabensammlung zur Funktionentheorie* (Berlin: Walter de Gruyter & Co., 1923), a collection of several hundred exercises on this material with their solutions, which, incidentally, will put little strain on the reader's German.

The physical appearance of the book and its translation are good; the price is moderate. Similar publication of several other volumes in the German "pocket book" series would constitute a real service to students of both statistics and pure mathematics.

Curve Fitting for Students of Economics. *Brij Narain* (Professor of Economics, Sanatana Dharma College, Lahore, India). Lahore, India: S. Chand & Co., 1944. Pp. viii, 107. Rs. 10/-.

Review by John H. Smith
Acting Chief Statistician, Bureau of Labor Statistics
Washington, D. C.

In the first four chapters Narain presents elementary applications of least squares to original data and to their logarithms. The fifth chapter is devoted to population concepts such as life tables, net and gross reproduction rates, and age composition. In the last two chapters is presented a method of

fitting curves to population data and to frequency data in ogive form to which "the attention of the reader is specially drawn." If x represents population, an arbitrary time origin is chosen at which $x=a/2$. The trend value of x at time t is determined from the polynominal trend value of K, the average rate of change in $\log x-\log(a-x)$, from the origin to t. The presentation is made more difficult to understand by the fact that the author calls a the "maximum population" when most of the examples show population for which the trend declines before reaching this "maximum." This method might be adapted to special types of problems. Even if less flexible trend functions were chosen for K, however, the method could not replace forecasts of population based on trends in birth and death rates and changing age composition of the population. Except for this specialized type of curve and the use of Indian data with accompanying discussion, this book has little in it which would be of interest to readers who are familiar with the much more comprehensive books by writers in England and the United States such as the ones by Croxton and Cowden, Ezekiel, Elderton, and others.

LETTERS ABOUT BOOKS

Readers are invited to submit letters about statistical methodology books for publication in this forum. Concise, informative letters which supplement previously published reviews by pointing out specific strengths, weaknesses, errors, and errata in currently used books are wanted. Criticisms based on actual use of a book as a text are especially desired from statistics instructors. Other letters may consist of suggestions for the writing of books and reviews. Letters which contain adverse criticisms of JOURNAL *reviews will be submitted to the author of the review for any reply he may care to make. Contributors are requested to avoid personalities. The right to decide whether a letter merits publication is reserved. Letters should be sent to the review editor, Oscar K. Buros, Rutgers University, New Brunswick, N. J.*

SEQUENTIAL ANALYSIS OF STATISTICAL DATA: APPLICATIONS

A REVIEW by B. L. Welch of *Sequential Analysis of Statistical Data: Applications*, contains the following sentence: "A basic criticism which has been made of Wald's work is that frequently m will be about halfway between m_1 and m_2, and that therefore appreciable average reductions in sampling size will not happen." I should like to comment briefly on this.

If it were true that a substantial part of a factory's output had to be characterized by m about halfway between m_1 and m_2, much more would be wrong with the sampling inspection scheme than merely failure to effect economies in inspection. The probability of acceptance at m is of the order $\frac{1}{2}$. (By this is meant that it is fairly large. Its exact value of course depends upon the various parameters involved.) Now if product characterized by m is satisfactory, than either the producer or the consumer has a legitimate grievance in that a substantial, satisfactory portion of the output is being rejected or completely inspected with the cost that this implies. If product characterized by m is defective, then a substantial fraction of the output is defective, and these defective items are being disposed of partly by acceptance by the consumer and partly by rejection (at the producer's expense, presumably). Such an economic setup cannot long endure and calls for changes more drastic than a change in sampling plans. It seems to me that in a production system which has reached at least the approximate stability where one can be concerned with such questions as minimizing average inspection, the bulk of the output must be in the

neighborhood of m_2; consequently, the criticism is invalid.

Now m_2 denotes good product and m_1 unsatisfactory product. Let the process average be denoted by m which therefore lies near m_2. A Wald sequential test minimizes inspection at m_1 and m_2. It seems to me that a next step should be to construct a test procedure which will have the prescribed power at m_1 and m_2 and will minimize inspection at the process average m. Using Wald's equation for the power of a sequential test such a test procedure can at present be constructed to a required degree of approximation only at large cost in computational effort. Clarification of this problem should be of service to those engaged in acceptance sampling and would no doubt contribute to the theory of Wald tests.

J. WOLFOWITZ, *Associate Professor, Institute of Statistics, University of North Carolina, Raleigh*

INDUSTRIAL STATISTICS

IN RESPONSE to your request for letters concerning recent books on statistical methodology we comment below on H. A. Freeman's *Industrial Statistics* (New York: John Wiley & Sons, Inc., 1942). We feel that the title and wide distribution of this book have caused many statistical novices to read it, and that it is therefore important to correct some of its more important errors. For many of these, the author should not be held responsible; they have been collected from the literature where they have the support of eminent statisticians.

THE USE OF THE t TEST. On page 17 Freeman points out that "a large value of t may reflect differences in variances rather than differences in means" and then passes to a test of homogeneity of variance. Why not include a simple, conservative version of the t test? One such is obtained by estimating the variance of each mean from the corresponding sample, adding these together to estimate the variance of the difference of means, and then using the smaller number of degrees of freedom. When the two samples are of equal size, the only change from the classical procedure is the halving of the number of degrees of freedom. Here one may point out that the "loss" of degrees of freedom is an example of the balance of sensitivity vs. security which occurs throughout all statistical procedures.

Immediately following the warning, Freeman analyzes an example where the variance ratio is 3.05, which is at about the 8 per cent point for 9 and 9 degrees of freedom (two-tailed F test). The lack of homogeneity is surely suspicious and might well be discussed, yet it is not. The 5 per cent point of t on a conservative 9 df would be only 2.262 instead of 2.101 on 18 df. This fact would help the reader or student who has to apply this technique in practice. We are reluctant to conclude from this warning and example that Freeman believes that whenever a test of significance fails to indicate different variances it is automatically safe to assume them the same, but a student might easily conclude this.

On page 62 we find "Had the entire set of grids differed significantly among themselves, the following procedure could have been used to determine whether or not the apparent best and second best grids differ significantly between themselves." After outlining the use of a t test after an F test, Freeman points out that the difference between the largest and the smallest will often appear "significant" due solely to error, and then warns the reader briefly that "A t test applied to two means after over-all homogeneity has either been refuted or not must be used with caution." He might also warn the reader about the case where the F test has not been made. Yet on the next page he uses the t test on the best and second best without comment! In noticing this point we do not wish to imply that a good solution exists—we know of none—but we feel that the reader should have been warned.

THE USE OF L_0 AND L_1 TESTS. On page 17 Freeman introduces the L_1 test, for application to the case of two variances. From a pedagogical point of view there seems much to be gained by using a two-tailed F test, which is perfectly equivalent and part of a widely applicable test. At the top of page 17 the sample sizes n_X and n_Y are potentially different, in the middle of page 17 the L_1 test is suggested, and it is not until page 87 that we learn that the author does not want to use

the L_1 test unless the sample sizes are equal. What is one supposed to do with two unequal samples? (Why not use the L_1 test?)

On page 87 Freeman states the questions which the L_0, L_1 and F tests are supposed to answer, but he makes no statement of the very important fact that these tests assume that the different samples came from normal populations whether or not the hypothesis tested is true. On page 91 these tests are then applied blindly to suspicious-looking data without inquiring as to normality.

CONCLUSIONS FROM TESTS OF SIGNIFICANCE. On page 24 we find that lack of statistical significance implies that "Hardness is really affected by stain, whereas breaking [bending] strength is not." That is, from lack of statistical significance we conclude lack of real effect—this seems highly unsound! In the actual situation, it is extremely probable that whatever chemical action is involved in staining does have some average effect on the bending strength. While the data probably do support the conclusion that the effect of stain on bonding strength is unimportant in practice, it is very important to distinguish between this conclusion and that stated on page 24.

There is a general tendency to discuss the case of two significance tests and a combined test in a confused or erroneous way. It is well known that if all tests are applied at the same level of significance the following possibilities arise: (*a*) all tests show significant effects; (*b*) one single test and the combined test show significant effects; (*c*) one single test shows a significant effect and no others do; (*d*) no single test shows a significant effect but the combined test does; and (*e*) no test shows a significant effect.

Case (*c*) arises infrequently, but, as the example $4+1=5$ for χ^2 on $1+1=2$ degrees of freedom shows at the 5% level, it can easily arise in practice. What conclusion to draw when it does arise is a difficult and, we believe, unsettled problem. (In fact, the general question of drawing conclusions when multiple tests are applied seems to be almost ignored by Freeman.) The discussion at the foot of page 61 is a case in point, although four tests not two are involved, for we are asked to conclude that if relatively insensitive tests for normality and for equality of variance detect nothing, then a large F value must imply differences in the means ("it follows that the part of H which is untenable is"). This neglects all possibilities of type (*c*)! On page 89 we have a similar situation—if the L_0 test is not significant, then the L_1 and F tests "must both fail (i.e., show no significance)."

Possibility (*d*) above is also hard to interpret; $3+3=6$ illustrates its occurrence for χ^2 on $1+1=2$ degrees of freedom. On page 90 we are told that such things cannot be, for if the L_0 test is significant and the L_1 test is not "hence they must differ in their means."

This treatment of overlapping tests of significance seems entirely unsatisfactory.

One of the most serious examples of drawing false conclusions from tests of significance occurrs in Freeman's discussion of regression problems, where significance of regression is misinterpreted to mean adequacy. On page 100 we are told that the "adequacy of a straight line . . . may be tested by determining the probability that . . . ." It is clear that this tests whether the slope of the regression line is different from zero and that it cannot, for example, indicate whether a curved regression line is needed. On page 102, "If the regression line is inadequate, the mean square due to regression will not be significantly larger than the residual or chance mean square." Chance mean square, indeed! this mean square contains all effects due to curvilinearity of regression! The same comment still applies. On page 104, "Linear regression does not account for a sufficient part of the total variability; it is not adequate for the purpose of prediction." This conclusion is reached on a total of 10 degrees of freedom and without consideration of the accuracy required! On pages 107–8 the question is asked "Would the regression . . . enable us effectively to predict warp breakage, or is the residual variability too great?" The question is not answered explicitly, but a regression on the "important influence, namely, relative humidity" is calculated, and found to be highly significant. It is not pointed out that the effect of regression is to reduce the mean square

from 1.23 to 1.18, and the standard error of estimate by less than 3 per cent. Is this effective prediction?

Finally on page 117 there is a combined conclusion, that *because* $F = 35.8$ the regression *describes the relationship between tensile strength, hardness and density.* There is no reason to believe that "this relationship," whatever that may mean, is linear.

THE PROBLEM OF NORMALITY. Freeman's attitude toward normality in small samples does not seem to be clear or consistent. On page 9 we are told that "One of the facts assumed to be known of the population . . . normally distributed" and this is followed on the same page by "The technique of testing the hypothesis that the population is normal is similar in general nature to the test of the hypothesis $\bar{d}' = 0$ . . . ." This is followed by a description of the a and b_1 tests of normality. These tests are applied to a sample of 15, with no specific indication of the wide variations from normality which would not be detected in such a sample. Freeman states only that "For small samples these tests are sensitive only to large departures from normality."

On page 21, we have a paired comparison of 28 differences, one of which is exceptionally large. This is clearly a fertile field for a test of normality, even with a sample as small as 28. Yet there is no such test, but only a discussion of the possible omission of this observation. Freeman states that omission is unsound without ancillary information, and does not discuss the very practical problem where apparently discrepant observations are due to nonnormality.

On page 25, we have a sample of 5 and a firm decision not to make a test of normality. We conclude that Freeman would stop testing somewhere between 15 and 5 observations. On page 27 we are told that if the sample is larger than 50, and if the population is as much as 10 times as large as the sample "the tendency to normality of the distribution of the means of random samples is negligibly affected by the nature of the population." The well-known Cauchy distribution and its many relatives show the incorrectness of this as a mathematical statement. Opinions may differ as to its application in practice, but we doubt its validity. The student may easily conclude that in samples of 50 or more it is useless to test normality, since it does not matter. We would hold, on the contrary, that samples of more than 50 are needed to make a test of normality which is sensitive enough to give any comfortable assurance of sufficient normality to warrant the use of the common tests at even the 5 per cent level. In practice, evidence about normality or its lack comes from broad experience or other samples.

On pages 37–38 Freeman develops a more or less conventional derivation of the normal distribution from many contributing effects. There is no indication that these effects *must* be additive. (Clearly if additive effects make X normally distributed, then multiplicative effects make e^x nonnormal!)

THE ANALYSIS OF VARIANCE. The introduction of the analysis of variance on page 58 seems exceedingly messy. We are unable to determine whether Freeman wishes to treat the column population means as fixed or as a random sample from a normal population. We feel that any adequate treatment must discuss and distinguish both approaches.

On page 63 we are told that "Five independent estimates of the variance of the population can be found, of which four are listed in the following table." There is no such fifth estimate.

On page 72 we are told that "The F test, as used in the analysis of variance, is essentially the ratio of variability associated with a suspected cause to error." This is, it seems to us, exceedingly liable to misinterpretation—only if the reader really understands the analysis of variance in advance can he interpret the word "associated" safely.

On the same page Freeman tests the variability of lot means. If he had wished to make the test on the assumption that either (a) the same nine rolls from each of the same three lots are to be tested and retested, or (b) there are exactly nine rolls in a lot; then it would have been proper to test against the mean square for error. If he wished to treat the rolls as samples of large lots of rolls, it would have been proper to test against the mean square for rolls within lots. The first is highly significant, the second very near the 50 per cent point. If there are

many rolls per lot the conclusion would be that there are differences from roll to roll of statistical significance and that these adequately explain the observed lot-to-lot variation. Freeman, however, tested against a pooled mean square for which we know of no reasonable interpretation!

The discussion of classification of interaction terms on page 85 seems unsatisfactory. It begins with the statement that "between pots" has no interest because "the pots were not used in any particular sequence." In their original paper (referred to by Freeman) Hampton and Gould state that, in this series of tests, the pots were selected to be as different as possible, the mean square found (the largest of all) suggests a definite systematic effect due perhaps to selection of pots or to the fact that the same two arches were used in all runs. Freeman states that "between runs" also means little because "the runs are quite independent of each other"—yet they were made at different times and contain the effects of weather and of the judgement of the furnace operator as to when the temperature should be reduced. The mean square for runs is the third largest in the analysis. Proceeding on these assumptions, which are likely to be incorrect, Freeman is then able, by some principle not clear to us, to classify together under "among cylinders" (page 86) terms with mean squares of 4566, 1922 and 183—by the size of their mean squares alone they are of vastly different nature. Freeman seems to imply that to interpret an interaction term it should be classified under one or the other main effects, a procedure which seems to us contrary to the basic principles of the analysis of variance.

No account is given of how the analysis of variance technique can be applied to the estimation of components of variance. This we have found at least as important as the significance test use of analysis of variance, and its omission we consider particularly unfortunate. The omission is the more surprising in that Tippett (upon whom the author has drawn for several examples) is very clear on the point.

On page 108, we are told that a certain subsidiary variance analysis may be useful "for linear regressions calculated from grouped data." Why only from grouped data? The statistician may group his data at least as intelligently as the man who groups it before letting the statistician see it.

On page 74 we have error piled on error. All the mean squares are tested against the mean square for error, where for the conclusions drawn both "cut" and "lot" should be tested against "interaction." Second, all the F critical values are incorrect—the degrees of freedom have been permuted in using the table. The correct critical value for "interaction vs. error" is about 2.31—the interaction is significant. The "lot vs. interaction" test is significant and nearly highly significant, while "cut vs. interaction" is far from significance. The correct conclusion is that there were significant lot differences and significant interaction of between lot and cut, but that there was no indication of an average effect of direction of cutting. Compare these with Freeman's conclusions.

On page 56 Freeman states that "it is unlikely . . . that just five machines would be used and that each would be used exactly once with each grid and each operator. If these conditions are not satisfied, machine effects cannot be removed." Nonorthogonal analysis of variance is not easy to analyze or interpret—but it is an essential tool in deciphering incomplete experiments.

REGRESSION. The treatment of regression is unsatisfactory in a number of respects. Some of these have been referred to elsewhere, in particular the use of a significance test as a criterion of "adequacy" of a regression equation.

Perhaps the most serious confusion exists on the question of the choice of independent variable. On page 113, it is stated that this choice "depends not on what we would like to predict, but on which of the two variables, X and Y, is free from error." The author goes on to say that if X is free from error, and we wish to predict X from Y, we should obtain the regression of Y on X, and then solve this for X. But on page 90, he treats a problem in which a measurement Y is accurate but costly and destructive, and it is desired to substitute a cheaper but less acurate nondestructive measurement X. To solve this problem he obtains the regression of Y on X.

This is not the place for a complete discussion of this problem; but it

should be pointed out that the statement on page 113 is wrong, and that the proper regression was fitted on page 99. It may be suspected that the statement on page 113 arose out of a misreading of a paper by Eisenhart to which reference is made. Eisenhart was concerned with the problem which arises when the values of one variable have been arbitrarily fixed. In this case, of course, only one regression line is available.

THE PHILOSOPHY OF EXPERIMENTATION. On page 5, Freeman indicates that it would have been better to omit "muck soil" from the pipe experiment if the experimenter knew in advance that it behaved differently from the other soils, because "its inclusion is uninformative and the loss in precision is costly." This conclusion is clearly not generally valid, for the real purpose of this experiment will often be to serve as a guide in choosing pipe rather than in estimating the average effect. In the situation envisaged here it would probably be wise to allot not 1, but 3 or 4 of the small number of pairs available to different mucky soils and to agree in advance that the mucky and nonmucky groups be analyzed separately.

On page 7, Freeman states that the intelligent experimenter will *always* use a method involving a variable rather than a method involving categories. Some attention must certainly be given to the relative difficulties of obtaining equal sized samples before an adequate answer can be given.

On page 67, we are told that the use of the Latin Square "wastes" degrees of freedom which could otherwise be allotted to error. This question arises not only for Latin Squares, but in every design situation where some degrees of freedom have been allotted to some variable of potential importance or effect. It seems to us that no waste is involved; if the mean square for this additional variable is about the size of the error, the average practitioner will, wisely or unwisely, pool these together with no loss of degrees of freedom, while if the mean square is large, it has been worthwhile by reducing the size of the error mean square.

BASIC ASSUMPTIONS. On page 44 we are told that the estimate of σ^2 based on the sum of squares of deviations from the sample mean is "best" in the sense of minimum variance. This occurs in a section discussing the unbiased nature of this estimate (Freeman introduces the excellent term "mean estimate" here), where every result except this one about minimum variance applies to arbitrary distributions. There is no warning that minimum variance is known only for the normal distribution, while for other distributions, e.g., the double exponential, other estimates have minimum variance. The same dangerous statement is applied to two samples at the top of page 47.

At the foot of page 98 we are told that if we draw chips (each bearing X and Y) from a bowl, and if the correlation between X and Y is zero, then the ordinary regression analysis leads to a quantity with the F distribution. This is, of course, only known to be true when the distribution of X and Y is bivariate normal, and is false when, for example, their distribution is bivariate uniform.

On page 106 we are told that "It is here assumed that the variances of all eight columns are equivalent, within the limits of chance variation. This assumption, which may be checked by the L_1 test, must be met for the F test to be valid." It seems to us that any interpretation that a student is likely to put on this is wrong. For example, these misinterpretations seem likely: (*a*) Unless the sample variances of the eight columns meet the L_1 test, the F test is invalid. (*b*) If the population variances of the eight columns are nearly enough equal to allow the sample to pass the L_1 test, the F test is exact. The proper interpretation, of course, would be: (*c*) The F test is only exact when the eight column populations are normal with the same variance; we may be able to detect large deviations from equality in the sample; the F test is not badly affected by lack of equality; we will not make too many serious mistakes if we use the F test, checking suspicious cases with the L_1 test and watering down our conclusions when the lack of homogeneity seems large.

CONTROL. On pages 131–32, we are introduced to a definition of control which seems to us widely different from that employed in practice. Implicitly it seems to state that a process is in control when its output is random. In Shewhart's original discus-

sion, he made it clear that he defined control economically, not statistically. Control meant that it was not *economic* to try to remove further causes of variation, which would almost certainly exist and cause non-randomness in any practical state of control. The control chart procedure was developed empirically to help to attain this state—and was to be used both before and after control was reached. Yet on page 133 we are told that the type of "systematic quality control" discussed in this chapter cannot be used until population homogeneity has been reached. On this internal evidence, then, these methods are only useful for the, possibly uneconomic, refinement of a quality control program set up by more useful methods.

In passing we note the *equality* between sample and population values which page 133 promises "will be shown" later. We presume that an averaging operation has been neglected.

The table on page 134 offered an alert expositor the chance to point out the change in level of defectives between the 6th and 7th days. Any plot of p would make this obvious. It seems poor pedagogy to point out that the results are not homogeneous, and *not* to take the data from the 7th day on and show how they behave. Noting the error of 9.8 for 7.8, six of the 7 values of "p . . . in following sample" on page 135 are below the control limits. The reader should be told how suspicious this is.

MINOR POINTS. In general there seem to be unduly many significant figures. On page 13, for a sample of 15, we are given an estimate of σ^2 with 6 "significant figures;" surely three could have been dropped without loss. But on page 117—we have a mean square on 2 *df* given to 10 "significant figures," where 3 could be regarded as slightly excessive. This is more interesting since, in the same table, a mean square on 57 *df* is given to three less "significant figures." This seems quite illogical.

The exponent of 4 instead of 1 for the denominator of the estimate of μ_4 on page 32 might be annoying. On page 69, the last parenthesis should be squared.

JOHN W. TUKEY, *Department of Mathematics Princeton University*

CHARLES P. WINSOR, *Department of Biostatistics, School of Hygiene and Public Health, Johns Hopkins University*

ELEMENTARY STATISTICAL METHODS

ELEMENTARY Statistical Methods as Applied to Business and Economic Data (New York: Macmillan Co., 1943) by William Addison Neiswanger is an elementary text designed especially for use in a first-semester course. The theory of statistics is illustrated with concrete examples. Formulas are given without proof. This is, of course, necessary because an economics freshman would not have enough mathematical training to follow the derivations of the formulas. But the author has gone too far in distorting the theory of statistics.

In regard to normal distribution, the author said (page 209), "This normal distribution is also referred to as 'the bell-shaped curve,' 'the curve of errors,' 'the binomial distribution,' all of which, while having technical differences as to meaning, are commonly used interchangeably when referring to this distribution." Since the author realizes that these terms have technical differences, he should discourage rather than encourage their misuse. It is especially absurd to call a normal distribution a binomial distribution. The author states repeatedly (page 249), "normal in the Gaussian sense." Does a normal distribution have any other sense?

In discussing the merits of the different kinds of averages such as the mean and the median, the author failed to mention the most important property of the mean, i.e., the mean generally is a more efficient statistic than the median.

The author called (p. 340) the coefficient of variation "a pure number which describes relative variation." It is true that this coefficient is independent of unit, but he did not mention the fact that the coefficient of variation depends on the origin which is an undesirable property of this coefficient. A mean can be zero and indeed can be a negative number.

The author defined (p. 346) the standard error of the mean as "the standard deviation of a series of means computed from large samples, drawn

at random from a single, homogeneous universe."

First, the size of the samples, large or small, has nothing to do with the definition of the standard error, but the samples must be of the same size. Second, the meaning of "a series of means" is not clear, but it was interpreted later (p. 349) as the means of "a large number of random samples." The number of the samples is quite irrelevant to the definition of the standard error of the means. The "series of means" must be composed of all possible means, the number of which may be large or small.

The author did not give the exact relation between the standard error of the mean and the standard deviation of the population which is $\sigma_M = \sigma/\sqrt{N}$. Instead, he gave an equation (p. 348) showing the relation of the standard error of the mean and the standard deviation of the sample without saying that it is only an approximate relation.

In testing the significance of the difference between two sample means, the author states the null hypothesis as this (p. 357): "Our problem here is to determine whether both might reasonably be viewed as having come from one and the same parent population or universe," but the standard error of the difference between the sample means is given as

$$\sigma_D = \sqrt{\frac{\sigma_{x_1}^2}{N_1 - 1} + \frac{\sigma_{x_2}^2}{N_2 - 1}}$$

where $\sigma_{x_1}^2$ and $\sigma_{x_2}^2$ are the variances of the samples and not the unbiased estimates of the population variance. If the null hypothesis is that both samples are drawn from the same population, a pooled estimate of the population variance should be used. This formula would be correct if the null hypothesis were that the samples are drawn from two different populations with the same mean. Throughout the discussion, the author failed to draw a distinction between the two different null hypotheses.

The presentation of "Analysis of Functional Relationship" is very awkward. The regression does not deal with the analysis of the functional relationship between two variables x and y, as the author suggested. It deals with the functional relationship between x and the means of the arrays of y.

The author spent 45 pages in discussing the linear regression, but he did not even give the definition of the line of regression. Intentionally or unintentionally the author gave the impression that the line fitted by the method of least squares is the line of regression. He did not say that the method of least squares is only a method of estimating the line of regression of the population. In fact, he cannot possibly make such a statement because the specific population is given at the end rather than at the beginning of the chapter on linear regression.

The population was specified as follows (p. 630): "The most important assumptions are four in number: (1) The assumption of linearity. (2) The assumption of normality. (3) The assumption of independent events. (4) The assumption that Y is a function of X." The important assumption of homoscedasticity is not included.

The assumption of linearity is erroneously explained as ". . . straight lines will describe the functional relation between the variables." The assumption of linearity is known to the writer as the locus of the means of the arrays of y being a straight line.

For the assumption of normality, the author gave a bivariate normal distribution (p. 643). This is a sufficient assumption, but the assumption that x is also normally distributed is not necessary.

Since the assumption of homoscedasticity is not included in the discussion, the meaning of the standard error of estimate has never been made clear. The residue sum of squares divided by $N-2$ is the unbiased estimate of the variance of the arrays of y. The author tried to present this fact (p. 634) by quoting a formula from Mordecai Ezekiel's *Methods of Correlation Analysis*, 1930, where the unbiased estimate is given as $S_y^2(N-1)/(N-2)$ but Ezekiel's S_y^2 is the residue sum of squares divided by $N-1$ (p. 114, Ezekiel), while the author's S_y^2 is the residue sum of squares divided by N. Therefore, the correction should be $N/(N-2)$ rather than $(N-1)/(N-2)$.

This can hardly be called a good book, not because it is worse than some of the existing elementary statistics books, but because it shows no improvement over them.

JEROME C. R. LI, *Instructor in Mathematics, Oregon State College*

PUBLICATIONS RECEIVED

Statistical methodology publications received are listed elsewhere in this issue under the section "Statistical Methodology Index."

Conference on Research in Income and Wealth. Studies in Income and Wealth, Vol. 8. New York 23: National Bureau of Economic Research Inc. (1819 Broadway), 1946. Pp. xiv, 297. $3.00.

Proceedings of the Conference on Quartermaster Textile Research. Published for the Inter-Society Council for Textile Research. New York 16: Textile Research Institute, Inc. (10 East 40th St.), 1945. Pp. vi, 61, $1.00. Paper.

Financing Business During the Transition. Charles C. Abbott. Committee for Economic Development Research Study. New York 18: McGraw-Hill Book Co., Inc. (330 W. 42nd St.), 1946. Pp. xii, 141. $1.75.

Rising State Expenditures. Project Note No. 15. New York: Tax Foundation (30 Rockefeller Plaza), 1946. Pp. 19. Paper, mimeographed.

Significant Trends in State Revenues. Project Note No. 16. New York: Tax Foundation (30 Rockefeller Plaza), 1946. Pp. 14. Paper, mimeographed.

The Regulation of the Security Markets. Willard E. Atkins, George W. Edwards, and Harold G. Moulton. Washington 6, D. C.: The Brookings Institution (722 Jackson Place), 1946. Pp. vi, 126. $2.00.

The Office Library of an Industrial Relations Executive, Fifth Edition. Helen Baker. Princeton, N. J.: Industrial Relations Section, Princeton University, 1946. Pp. vi, 35. $0.50. Paper.

Circular and Hyperbolic Functions: Exponential and Sine and Cosine Integrals, Factorial Functions & Allied Functions, Hermitian Probability Functions, Second Edition. British Association for the Advancement of Science, Mathematical Tables, Vol. 1, Prepared by the Committee for the Calculation of Mathematical Tables. London: Cambridge University Press, 1946. Pp. xi, 72. 10s. (New York 11: Macmillan Co. (60 Fifth Ave.). $2.50.)

Statistical Abstract of the United States, 1944–45: Sixty-Sixth Number. Compiled by the Bureau of the Census. Washington 25: Government Printing Office, 1945. Pp. xlii, 1023. $1.75. *Reviewed in this issue.*

Economic Research and the Keynesian Thinking of Our Times. Arthur F. Burns. Twenty-Sixth Annual Report of the National Bureau of Economic Research. New York 23: National Bureau of Economic Research, Inc. (1810 Broadway), 1946. Pp. v, 69. Paper.

An Introduction to Modern Economics. Valdemar Carlson. Philadelphia 5, Pa.: The Blakiston Co. (1012 Walnut St.), 1946. Pp. xvii, 337. $3.50.

The Theory of Monopolistic Competition: A Re-orientation of the Theory of Value, Fifth Edition. Edward Hastings Chamberlin. Harvard Economic Studies, Vol. 38. Cambridge 38, Mass.: Harvard University Press, 1946. Pp. xiv, 282. $3.50.

Population in Modern China. Ta Chen. Chicago 37, Ill.: The University of Chicago Press (5750 Ellis Ave.), 1946. Pp. ix, 126. $2.50.

Disability Among Gainfully Occupied Persons: An Introduction to Disability Insurance Statistics. I. S. Falk, Barkev S. Sanders, and David Federman. Bureau Memorandum No. 61. Washington, D. C.: Bureau of Research and Statistics, Social Security Board, Federal Security Agency, 1945. Pp. iv, 60. Paper.

Monetary Policies and Full Employment. William Fellner. Publications of the Bureau of Business and Economic Research, University of California. Berkeley 4, Calif.: University of California Press, 1946. Pp. xx, 268. $3.50.

American Foundations for Social Welfare: Includes a Directory of 505 Foundations. Shelby M. Harrison, and F. Emerson Andrews. New York 10: Russell Sage Foundation (130 E. 22nd St.), 1946. Pp. 1, 249. $2.00.

Government Statistics for Business Use. Edited by Philip M. Hauser (Assistant Director, Bureau of the Census) and William R. Leonard (Deputy Chief, Division of Statistical Standards, Bureau of the Budget). New York 16: John Wiley & Sons, Inc. (440 Fourth Ave.), 1946. Pp. vi, 432. $5.00. *To be reviewed.*

California Business Cycles. Frank L. Kidner. Publications of the Bureau of Business and Economic Research, University of California. Berkeley 4, Calif.: University of California Press, 1946. Pp. xiv, 131. $2.50.

Theory of Functions: Part 1, Elements of

the General Theory of Analytic Functions. Konrad Knopp. Translated from the fifth German edition by Frederick Bagemihl. New York 19: Dover Publications (1780 Broadway), 1945. Pp. vii, 140. $1.25. *Reviewed in this issue.*

Full Production Without War. Harold Loeb. Princeton, N. J.: Princeton University Press, 1946. Pp. xviii, 284. $3.50.

Controlling World Trade: Cartels and Commodity Agreements. Edward S. Mason. Committee for Economic Development Research Study. New York 18: McGraw-Hill Book Co., Inc. (330 W. 42nd St.), 1946. Pp. xvii, 289. $2.50.

Changes in Income Distribution During the Great Depression. Horst Mendershausen. Conference on Research in Income and Wealth, Studies in Income and Wealth, Vol. 7. New York 23: National Bureau of Economic Research, Inc. (1819 Broadway), 1946. Pp. xviii, 106. $2.50. *To be reviewed.*

Production Credit for Southern Cotton Growers. A. E. Nielsen. New York 27: King's Crown Press (2960 Broadway), 1946. Pp. ix, 104. $2.50. Paper.

Economic Factors Influencing Railroad Employment. Maurice Parmelee. Chicago 11, Ill.: United States Railroad Retirement Board (844 Rush St.), 1946. Pp. viii, 152. Paper, mimeographed.

A Review of Railway Operations in 1945. Julius H. Parmelee. Special Series No. 75. Washington 6, D. C.: Association of American Railroads (Transportation Bldg.,) 1946. Pp. 32. Paper.

A Trial on Trial: The Great Sedition Trial of 1944. Maximilian St. George, and Lawrence Dennis. Chicago, Ill.: M. J. St. George (10 S. La Salle St.), 1946. Pp. 503. $5.00.

Personnel Problems Affecting Social Scientists in the National Civil Service. Lloyd M. Short. New York 17: Social Science Research Council (230 Park Ave.), April 1946. Pp. iv, 36. Paper.

Some European Currency and Exchange Experiences: 1943–1946. Frank A. Southard, Jr. Essays in International Finance, No. 7. Princeton, N. J.: International Finance Section, Princeton University, 1946. Pp. 23. Paper.

Domestic Servants in the United States, 1900–1940. George J. Stigler. Occasional Paper 24. New York 23: National Bureau of Economic Research, Inc. (1819 Broadway), April 1946. Pp. iv, 44. $0.50. Paper.

Economic Stagnation or Progress: A Critique of Recent Doctrines on the Mature Economy, Oversavings, and Deficit Spending. Ernst W. Swanson, and Emerson P. Schmidt. New York 18: McGraw-Hill Book Co., Inc. (330 W 42nd St.), 1946. Pp. xi, 212. $2.50.

International Economic Cooperation. J. Tinbergen. New York 3: Elsevier Publishing Co., Inc. (215 Fourth Ave.), 1945. Pp. 208. $4.00.

STATISTICAL METHODOLOGY INDEX, NO. 5

A QUARTERLY GUIDE TO CURRENT LITERATURE

Edited by

OSCAR KRISEN BUROS

Rutgers University

This bibliographic service presents statistical methodology literature (articles, books, theses, and chapters) published in 1945 to date. Anonymous references are listed first; other references follow in alphabetical sequence by first-named authors. Volume number, issue number (in parentheses), pagination, and date of issue are given for each journal reference. Stars indicate books, theses, and pamphlets; asterisks indicate publications seen by the editor of this INDEX. *Authors of papers on statistical methodology are invited to send reprints to Oscar K. Buros, Rutgers University, New Brunswick, New Jersey in order to facilitate prompt and accurate listings. Foreign-language papers should be accompanied by English translations of title. Information is desired about references omitted.*

The Indian Statistical Institute. *Nature* 156 (3072):722 D 15 '45.* [548

Missing-Plot Techniques. R. L. Anderson (Associate Professor of Experimental Statistics and Agricultural Economics, Institute of Statistics, University of North Carolina, Raleigh). *Biometrics B* 2(3):41–7 Je '46.* [549

Foundry Applications of Quality Control. William M. Armstrong (Ontario Research Foundation, 43 Queen's Park, Toronto 5, Canada). *Ind Qual Control* 2(6):12–6 My '46.* [550

Statistical Control in Manufacturing Steel Strips and Tin Plate. W. M. Armstrong. *Iron & Steel Eng* 23(7):69–77 Jl '46.* [551

Linear "Curves of Best Fit." Letter. A. E. W. Austen and H. Pelzer. (E.R.A. Laboratory, 5 Wadsworth Road, Perivale, Greenford, Middlesex, England.) *Nature* 157 (3995):694 My 25 '46.* [552

Index Numbers and Analysis of Gross Profit Variation as a Management Tool. Charles F. Axelson (Assistant Controller, United States Gypsum Co., Chicago). *J Account* 81(6):463–5 Je '46.* [553

Practical Statistics for Practical Physicists (Continued). Ralph Hoyt Bacon (Fairchild Camera and Instrument Corp., Jamaica, N. Y.). *Am J Physics* 14(3):198–200 My-Je '46.* [554

Table of Percentage Points of the t-Distribution. Elizabeth M. Baldwin (Post Office Research Station, Dollis Hill, London, N.W.2, England). *Biometrika* 33(4):362 Je '46.* [555

Measurement of Affective Power in Terms of Ratios of Partial ϵ^2 S. J. G. Beebe-Center and Roxanna Beebe-Center. (Harvard University.) *Am J Psychol* 59(2):290–5 Ap '46.* [556

Limitations of the Application of Fourfold Table Analysis to Hospital Data. Joseph Berkson (Head, Section of Biometry and Medical Statistics, Mayo Clinic, Rochester, Minn.). *Biometrics B* 2(3):47–53 Je '46.* [557

Finite Non-homogeneous Chains. David Blackwell (Assistant Professor of Mathematics, Howard University). *Ann Math* 46(4):594–9 O '45.* [558

An Experimental Design for Slope-Ratio Assays. C. I. Bliss (Biometrician, Connecticut Agricultural Experiment Station; Lecturer in Biometry, Yale University). *Ann Math Stat* 17(2):232–7 Je '47.* [559

Computation of Factors for Tolerance Limits on a Normal Distribution When the Sample Is Large. Albert H. Bowker (Assistant Director, Statistical Research Group Columbia University). *Ann Math Stat* 17(2):238–40 Je '46.* [560

Estimation of Genes in Inheritance of Quantitative Characters. Letter. S. W. Boyce (Wheat Research Institute, Christchurch, New Zealand). *Nature* 157(3994):657 My 18 '46.* [561

★Industrial Experimentation. K. A. Brownlee (Research Department, Distillers Co., Ltd., Great Burgh, Epsom, Surrey, England). A revision of the 1945 memorandum of same title reviewed in this JOURNAL, March 1946. Directorate of Ordnance Factories (Explosives), Ministry of Supply.

London, W.C.2: H. M. Stationery Office, 1946. 2s. Pp. 110. Paper.* (New York 20: British Information Service, 30 Rockefeller Plaza, 1946. $0.60.*) *To be reviewed.* [562

The Use of 2×n Chi Square in the Analysis of Change in Age Distributions. Paul T. Bruyere (Statistician) Martha C. Bruyere (Statistician), and Geraldine A. Gleeson. (Division of Venereal Diseases, U. S. Public Health Service, Bethesda, Md.) *Am J Pub Health* 36(5):510–4 My '46.* [563

★**Measuring Business Cycles.** Arthur F. Burns (Director of Research, 1946–; Professor of Economics) and Wesley C. Mitchell (Director of Research, 1920–1946; Professor Emeritus of Economics). (National Bureau of Economic Research; Columbia University.) Studies in Business Cycles, No. 2. New York 23: National Bureau of Economic Research, Inc. (1819 Broadway), 1946. Pp. xxvii, 560. $5.00.* *To be reviewed.* [564

★**How to Read Statistics.** R. L. C. Butsch (Formerly Professor of Education, Marquette University). Milwaukee 1, Wis.: Bruce Publishing Co. (540 N. Milwaukee St.), 1946. Pp. vii, 184. $2.50.* *To be reviewed.* [565

The Effect on a Distribution Function of Small Changes in the Population Function. Burton H. Camp (Professor of Mathematics, Wesleyan University). *Ann Math Stat* 17(2):226–31 Je '46.* [566

Remarks on Induction and Truth: A Symposium on Probability, Part III. Rudolf Carnap (Professor of Philosophy, University of Chicago). *Philos & Phenom Res* 6(4):590–602 Je '46.* *See also* 606, 614, 615, 640. [567

Statistical Techniques in Applied Psychology. E. G. Chambers (Psychological Laboratory, Cambridge University). *Biometrika* 33(4):269–73 Je '46.* [568

Information Given by Odd Moments. Edmund Churchill (Instructor in Mathematics, Rutgers University). *Ann Math Stat* 17(2):244–6 Je '46.* [569

Most Economic Sampling for Chemical Analysis. C. West Churchman (Assistant Professor of Philosophy, University of Pennsylvania). *Ind & Eng Chem, Anal Ed* 18(4):267–8 Ap '46.* [570

Relative Accuracy of Systematic and Stratified Random Samples for a Certain Class of Populations. W. G. Cochran (Associate Director, Institute of Statistics, University of North Carolina, Raleigh). *Ann Math Stat* 17(2):164–77 Je '46.* [571

★**Rudimentary Mathematics for Economists and Statisticians.** W. L. Crum (Professor of Economics) and Joseph A. Schumpeter (Professor of Economics). (Harvard University.) A revision of Supplement to *Quarterly Journal of Economics*, Vol. 52, May 1938. New York 18: McGraw-Hill Book Co., Inc. (330 West 42nd St.), 1946. Pp. xi, 183. $2.50. *To be reviewed.* [572

Mathematical Probabilities in Games of Chance: The Game of Sevens. E. Bonavent Dagabert (San Francisco, Calif.). *Math Teach* 39(4):155–8 Ap '46.* [573

Aid to Control of Open Hearth Practice by Statistical Methods. R. E. Dienst (Production and Cost Analyst, Great Lakes Steel Corp., Ecorse, Detroit, Mich.). *Iron & Steel Eng* 23(5):68–74, discussion 74–8 My '46.* [574

Statistical Quality Control. M. Dolan. *Eng Inspection* 11(1):11–2 sp '46.* [575

★**Statistical Analysis for Students in Psychology and Education.** Allen L. Edwards (Associate Professor of Psychology, University of Washington). New York 16: Rinehart & Co., Inc. (232 Madison Ave.), 1946. Pp. xviii, 360. $3.50. *Reviewed in this issue.* [576

The Future of Quality Control. George D. Edwards (Director of Quality Assurance, Bell Telephone Laboratories, Inc., New York). *Ind Qual Control* 2(6):10–9 My '46.* [577

Quality Control—A Management Tool. M. Herbert Eisenhart (President, Bausch & Lomb Optical Co., Rochester, N. Y.). *Ind Qual Control* 2(6):5–7 My '46.* [578

Mathematics Eases Tolerances Safely. William S. Everett (Lt.-Com., USNR, Office of Inspector of Naval Material). *SAE J* 54(4):28+ Ap '46.* [579

The Law of the Iterated Logarithm for Identically Distributed Random Variables. Abstract. Will Feller (Professor of Mathematics, Cornell University). *B Am Math Soc* 52(5):430 My '46.* [580

A Limit Theorem for Random Variables With Infinite Moments. W. Feller. *Am J Math* 68(2):257–62 Ap '46.* Abstract: *B Am Math Soc* 52(5):430 My '46.* [581

The Disproportionate Method of Market Sampling. Robert Ferber (Formerly Economist, Industrial Surveys Co., Chicago). *J Bus* 19(2):67–75 Ap '46.* [582

Operating Characteristics for the Common Statistical Tests of Significance. Charles D. Ferris (Engineering Statistician), Frank E. Grubbs (Chief, Surveillance Branch), and Chalmers L. Weaver. (Ballistic Research Laboratory, Aberdeen Proving Ground, Md.) *Ann Math Stat* 17(2):178–97 Je '46.* [583

¶**The Significance of Difference Between Means Without Reference to the Frequency**

Distributive Function. Leon Festinger (Research Center for Group Dynamics, Massachusetts Institute of Technology). *Psychometrika* 11(2):97–105 Je '46.* [584

Orthogonal Partitions of the 5×5 Latin Squares. D. J. Finney (Lecturer in the Design and Analysis of Scientific Experiment, University of Oxford). *Ann Eug* 13(1):1–3 Ap '46.* [585

Standard Errors of Yields Adjusted for Regression on an Independent Measurement. D. J. Finney. *Biometrics B* 2(3):53–5 Je '46.* [586

Statistical Analysis of Spontaneous Electrical Fluctuations. Letter. R. Fürth (Department of Mathematical Physics, University of Edinburgh) and D. K. C. MacDonald (Military College of Science, Blurton, Longton, Stoke-on-Trent, England). *Nature* 157(3998):807 Je 15 '46.* [587

Contributions to the Theory of Sequential Analysis: Part I. M. A. Girshick (Principal Statistician, Bureau of Agricultural Economics, Washington). *Ann Math Stat* 17(2):123–43 Je '46.* [588

How Quality Engineering Helps the Purchasing Agent. Eugene Goddess (Head of Factory Engineering, North American Philips Co., Dobbs Ferry, N. Y.). *Purchasing* 20(3):88–91 Mr '46.* [589

Tolerances in Quality Control. Eugene Goddess. *Mill & Fact* 39(1):110–3 Jl '46.* [590

Measures of Fluctuation of a Variable Mean. Abstract. Casper Goffman (Staff Assistant, Quality Control Department, Westinghouse Electric & Mfg. Co., East Pittsburgh, Pa.). *B Am Math Soc* 52(5):439–40 My '46.* [591

Use of Statistics in Writing Specifications. Casper Goffman (Staff Assistant) and Joseph Manuele (Director). (Quality Control Department, Westinghouse Electric Corp., East Pittsburgh, Pa.) *ASTM B* (130):13–7 Mr '46.* [592

Approximate Formulas for the Percentage Points and Normalization of t and χ^2. Henry Goldberg and Harriet Levine. (Statistical Research Group, Columbia University.) *Ann Math Stat* 17(2):216–25 Je '46.* [593

Statistical Methods in Cereal Chemistry. C. H. Goulden (Officer-in-Charge) and Allan E. Paull. (Dominion Laboratory of Cereal Breeding, Winnipeg, Manitoba, Canada.) *Biometrics B* 2(2):26–30 Ap '46.* [594

New Statistical Criteria for Learning and Problem Solution in Experiments Involving Repeated Trials. David A. Grant (Associate Professor of Psychology, University of Wisconsin). *Psychol B* 43(3):272–82 My '46.* [595

On the Accumulation of Chance Effects and the Gaussian Frequency Distribution. Letter. A. H. R. Grimsey (Military College of Science, Blurton, Longton, Stoke-on-Trent, England). *Philos Mag* 36(255). 294–5 Ap '45.* [596

An Approach for Quantifying Paired Comparisons and Rank Order. Louis Guttman (Associate Professor of Sociology, Cornell University). *Ann Math Stat* 17(2):144–63 Je '46.* [597

The Test-Retest Reliability of Qualitative Data. Louis Guttman. *Psychometrika* 11(2):81–95 Je '46.* [598

Limits for Control Charts. Preston C. Hammer (Partner, Probity Engineers, 5448 Maplehurst Ave., Los Angeles 41, Calif.). *Ind Qual Control* 2(6):20–2 My '46.* [599

The Covariance Analysis of Multiple Classification Tables With Unequal Subclass Numbers. L. N. Hazel (Associate Animal Husbandman, Western Sheep Breeding Laboratory, Dubois, Idaho). *Biometrics B* 2(2):21–5 Ap '46.* [600

Statistical Control of Accidents. J. M. Howell (Quality Control Engineer) and Lee Johnson (Safety Engineer). (Northrop Aircraft Inc., Hawthorne, Calif.) *Iron Age* 158(3):56–9 Jl 18 '46.* [601

On the Correlation Between a Variable and Its Super-Factor. E. H. Hsü (Department of Psychology, Catholic University of America). *J Psychol* 22(1):89–92 Jl '46.* [602

A Quick Method for Multiple R and Partial r's. William Leroy Jenkins (Associate Professor of Psychology, Lehigh University). *Ed & Psychol Meas* 6(2):273–86 su '46.* [603

A Useful Method for the Routine Estimation of Dispersion From Large Samples. A. E. Jones (Rothamsted Experimental Station, Harpenden, Herts, England). *Biometrika* 33(4):274–82 Je '46.* [604

On the Average of a Certain Wiener Functional and a Related Limit Theorem in Calculus of Probability. M. Kac (Assistant Professor of Mathematics, Cornell University). *Trans Am Math Soc* 59(3):401–14 My '46.* [605

On the Nature of Inductive Inference: A Symposium on Probability, Part III. Felix Kaufmann (New School for Social Research). *Philos & Phenom Res* 6(4):602–9 Je '46.* **Rejoinder to Mr. Kaufmann's Reply.** Rudolf Carnap (Professor of Philosophy, University of Chicago). *Philos & Phenom Res* 6(4):609–11 Je 46.* *See also* 567, 614, 615, 649. [606

★A Teoria da Indução Estatística. [The Theory of Statistical Induction.] Jorge Kingston (Professor, University of Brazil). Rio de Janeiro, Brazil: Serviço Gráfico do Instituto Brasileiro de Geografia e Estatística, 1945. Pp. 121. Paper. *To be reviewed.* [007

On the Solution of Some Equations in Least Squares. D. V. Lindley. *Biometrika* 33(4):326–7 Je '46.* [008

Improvement of Precision by Repeated Measurements: Application to Analytical Control Methods. John Mandel (The BG Corp., 136 West 52nd St., New York). *Ind & Eng Chem, Anal Ed* 18(5):280–4 My 15 '46.* [009

A Complete Quality Control Program. Joseph Manuele (Director of Quality Control, Westinghouse Electric & Mfg. Co., East Pittsburgh, Pa.). *Ind Qual Control* 2(6):9–12 My '46.* [010

A New Approach to Statistical Quality Control. Joseph Manuele. *Auto & Aviation Ind* 95(1):21+Jl 1 '46.* [011

Note on a Comment on the 'Correction' of Reliability Coefficients for Restriction of Range. Octavio A. L. Martins (Departamento Regional do Senai, Rio de Janeiro, Brazil). *J Ed Psychol* 37(3):182–3 Mr '46.* [012

Utility of Log-Log vs Arithmetic Coordinate Grids. A. A. Merrill (Motor Division, Industrial Divisions, General Electric Co.). *Gen Elec* 49(7):20–2 Jl 46.* [013

Comments on Donald Williams' Reply: A Symposium on Probability, Part III. Richard von Mises (Professor of Aerodynamics and Applied Mathematics, Harvard University). *Philos & Phenom Res* 6(4):611–3 Je '46.* Reply to 407; *see also* 507, 606, 615, 640. [614

Is the Laplacean Theory of Probability Tenable? A Symposium on Probability, Part III. Ernest Nagel (Associate Professor of Philosophy, Columbia University). *Philos & Phenom Res* 6(4):614–8 Je '46.* Reply to 407; *see also* 507, 606, 614, 640. [615

Statistical Control in Workshop Administration. W. T. O'Dea (Deputy Controller, London and South Eastern Region, Ministry of Production). *Inst Mech Eng Proc* 154(1):1–8, discussion 9–11 Je '46.* [616

Average Duration of Transition in Markoff Chains. Abstract. Isaac Opatowski. *B Am Math Soc* 52(5):440 My '46.* [617

Markoff Chains With Variable Intensities: Average Duration of Transition. Abstract. Isaac Opatowski. *B Am Math Soc* 52(5):440 My '46.* [618

Industrial Statistics Looks at the Future. Paul Peach (Industrial Statistician, Institute of Statistics, University of North Carolina, Raleigh). *Ind Qual Control* 2(6): 7–9 My '46.* [619

A New Method for Analyzing Aesthetic Preferences: Some Theoretical Considerations. E. A. Peel (Institute of Education, University of London). *Psychometrika* 11(2):129–37 Je '46.* [620

The Design of Optimum Multifactorial Experiments. R. L. Plackett and J. P. Burman. *Biometrika* 33(4):305–25 Je '46.* [621

Some Generalizations in the Multifactorial Design. R. L. Plackett. *Biometrika* 33(4):328–32 Je '46.* [622

On the Mean Conserving Property. C. Radhakrishna Rao (Statistical Laboratory, Presidency College, Calcutta, India). *Proc Indian Acad Sci, Sect A* 23(4):165–73 Ap '46.* [623

Statistical Control Applied to Financial Statements. William B. Rice (Director, Department of Statistics and Reports, Plomb Tool Co., Los Angeles, Calif.). *Acctg R* 21(3):267–72 Jl '46.* [624

Mathematics in Government Service and Industry: Some Deductions From the War-Time Experience of the Admiralty Computing Service. D. H. Sadler and John Todd. *Nature* 157(3992):571–3 My 4 '46.* [625

Probability Problems in Nuclear Chemistry. Erwin Schrodinger (Dublin Institute for Advanced Studies, Dublin, Ireland). *Proc Royal Irish Acad, Sect A* 51(1):1–8 D '45.* [626

The Low Cost of High Quality. F. C. Schulze (Chief Inspector, Hunter Pressed Steel Co., Lansdale, Pa.). *Mill & Fact* 38(3,4):100–14, 105–10 Mr, Ap '46.* [627

Dollars for Your Thoughts. Leslie E. Simon (Colonel, Ordnance Dept., Aberdeen Proving Ground, Md.). *ASTM B* (139):17–21 Mr '46.* [628

Asymptotic Distribution of Sums of Rademacher Functions. Jerome C. Smith (Department of Mathematics, Cornell University). *B Am Math Soc* 51(12):941–4 D '45.* [629

Quality Control in Production. Statistician. *Beama J* 53(104):42–4 F '46. [630

How Chemical Engineers Can Use Statistics. E. I. Stearns (Research Department, Calco Chemical Division, American Cyanamid Co., Bound Brook, N. J.). *Chem & Metal Eng* 53(5):110 My '46.* [631

On the Theory of Scales of Measurement. S. S. Stevens (Director, Psycho-Acoustic Laboratory, Harvard University). *Sci* 103(2684):677–80 Je 7 '46.* [632

The Analysis of Market Demand. Richard Stone (Director, Department of Applied Economics, University of Cambridge). *J Royal Stat Soc* 108(3–4):286–382, discussion 383–91 pts 3–4 '45.* [633

Bias in the Use of Small-Size Plots in Sample Surveys for Yield. P. V. Sukhatme (*Statistical Adviser, Imperial Council of* Agricultural Research, New Delhi, India). *Cur Sci* 15(5):119–20 My '46.* Letter: *Nature* 157(3993):630 My 11 '46.* [634

Random Sampling for Estimating Rice Yield in Kolaba, Bombay. P. V. Sukhatme; B. S. Kadam, and N. R. Bhat (Department of Agriculture, Bombay, India). *Proc Indian Acad Sci, Sect B* 23(4):194–209 Ap '46.* [635

Selected References on Statistics, the Theory of Test Construction, and Factor Analysis. Frances Swineford (Research Associate, *Department of Education*) and Karl J. Holzinger (Professor of Education). (University of Chicago.) *Sch R* 54(6):364–7 Je '46.* [636

Mathematics as a Tool in the Water-Works Laboratory. Harold A. Thomas, Jr. (Assistant Professor of Sanitary Engineering, Harvard University). *J New England Water Works Assn* 60(2):153–61 Je '46.* [637

Tables for Testing the Homogeneity of a Set of Estimated Variances. Computed by Catherine M. Thompson and Maxine Merrington. *Prefatory note* (pp. 296–301) by H. O. Hartley (Scientific Computing Service, Ltd., 23 Bedford Sq., London W.C.1, England) and E. S. Pearson (Department of Statistics, University College, London W.C.1). *Biometrika* 33(4):296–304 Je '46.* [638

A Single Plane Method of Rotation. L. L. Thurstone (Professor of Psychology, University of Chicago). *Psychometrika* 11(2):71–9 Je '46.* [639

General Solution of the Analysis of Variance and Covariance in the Case of Unequal or Disproportionate Numbers of Observations in the Subclasses. Foi Tsao (National Central University, Chungking, China). *Psychometrika* 11(2):107–28 Je '46.* [640

A Normalized Graphic Method of Item Analysis. William W. Turnbull (Head, Test Construction Department, College Entrance Examination Board, Princeton, N. J.). *J Ed Psychol* 37(3):129–41 Mr '46.* [641

Quality Control Cuts Costs. James Turner. *Am Bus* 16(6):14–5 Je '46.* [642

Minimal Variance and Its Relation to Efficient Moment Tests. J. R. Vatnsdal (*Associate Professor of Mathematics*, State College of Washington). *Ann Math Stat* 17(2):198–207 Je '46.* [643

The Probability Distribution of the Measure of a Random Linear Set. David F. Votaw, Jr. (Naval Ordnance Laboratory). *Ann Math Stat* 17(2):240–4 Je '46.* [644

Incomplete-Block Design Adapted to Paired Tests of Mosquito Repellents. F. M. Wadley (Statistical Consultant, Bureau of Entomology and Plant Quarantine, U. S. Department of Agriculture, Washington). *Biometrics B* 2(2):30–1 Ap '46.* [645

Tolerance Limits for a Normal Distribution. A. Wald (Professor of Mathematical Statistics) and J. Wolfowitz (Associate Professor of Mathematical Statistics). (Columbia University.) *Ann Math Stat* 17(2):208–15 Je '46.* [646

Some Order Statistic Distributions for Samples of Size Four. John E. Walsh (Department of Mathematics, Princeton University). *Ann Math Stat* 17(2):246–8 Je '46.* [647

Yule's "Characteristic" and the "Index of Diversity." Letter. C. B. Williams (*Rothamsted Experimental Station*, Harpenden, Herts, England). *Nature* 157(3989):482 Ap 13 '46.* [648

The Problem of Probability: A Symposium on Probability, Part III. Donald Williams (Associate Professor of Philosophy, Harvard University). *Philos & Phenom Res* 6(4):619–22 Je '46.* Reply to 614 and 615; *see also* 567, 606. [649

A Simple Test of Significance. Letter. E. J. Williams (P. O. Box 18, South Melbourne, Victoria, Australia). *Eng* 161(4193):496 My 24 '46.* A criticism of 407. [650

Inequalities in Terms of Mean Range. C. B. Winsten. *Biometrika* 33(4):283–95 Je '46.* [651

Should Engineers Study Statistics? Holbrook Working (Professor of Prices and Statistics, Stanford University). *J Eng Ed* 36(9):557–64 My '46.* [652

Periodicals Represented for the First Time

Am Bus—American Business. 12 issues; $3 (35¢); Dartnell Publications, Inc., 4660 Ravenswood Ave., Chicago 40, Ill.

Am J Pub Health—American Journal of Public Health. 12 issues; $5(50¢); 1790 Broadway, New York 19, N. Y.

Auto & Aviation Ind—Automotive and Aviation Industries. 24 issues in 2 vols.;

$1(25¢); Chilton Co., Chestnut & 56th Sts., Philadelphia 39, Pa.

Beama J—*Beama Journal.* 12 issues; (1s.); British Electrical & Allied Manufacturers' Association (Inc.), 36 Kingsway, London, W.C.2, England.

Ed & Psychol Meas—*Educational and Psychological Measurement.* 4 issues; $4($1.25); 917 15th St., N.W., Washington 5, D. C.

J Account—*The Journal of Accountancy.* 12 issues in 2 vols.; $4; 13 East 41st St., New York, N. Y.

J New England Water Works Assn—*Journal of the New England Water Works Association.* 4 issues; $4($1.25); 600 Statler Bldg., Boston, Mass.

Proc Indian Acad Sci, Sect B—*Proceedings of the Indian Academy of Science, Section B.* 4 issues; Rs. 18 (Rs. 2 or 3s); Indian Academy of Sciences, Bangalore, India.

Proc Royal Irish Academy, Sect A—*Proceedings of the Royal Irish Academy of Sciences.*

SAE J—*SAE Journal.* 12 issues; $10($1); Society of Automotive Engineers, 29 West 39th St., New York 18, N. Y.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 236 DECEMBER 1946 *Volume 41*

APPLICATIONS AND PROBLEMS OF PRODUCTIVITY DATA

CHARLES E. YOUNG
Westinghouse Electric Corporation

The gain in output per man-hour in manufacturing since 1899 appears to have been related fairly closely to increased use of power per worker. This raises the question whether the common assumption of a persistent rate of gain in productivity can be justified apart from corresponding increases in manufacturing investment per employed worker.

Manufacturing output has corresponded roughly to the product of an installed-horsepower index and the length of the work week. A sizable part of peak wartime output apparently came from overtime and extra-shift use of facilities, and apparently sustained high output and productivity in the early postwar years will require similarly intensive use of facilities.

Real hourly wages in manufacturing since 1914 have corresponded closely to changes in productivity, and finished goods prices have corresponded closely to changes in unit labor cost (wage rates adjusted for productivity).

AT LEAST two major developments have brought the question of productivity increasingly to the fore in recent years. One has been the popular statistical pastime of projecting gross national product from the starting assumption of full employment; the other has been in connection with disputes as to the justification of wage increases.

Without due consideration of the level of productivity, it is, of course, impossible to establish the quantity of output which is equivalent to stated assumptions as to the number of workers and their hours of work; it is also impossible to determine the quantity of output which employers, and hence customers, will receive for the funds expended in payrolls.

It is not my purpose in this paper to establish any innovations in technique, but simply to refresh your memories with some of the applications and problems of productivity data from the standpoint of a practicing business statistician.

Several aspects of the typical discussions of productivity are worthy of attention. First of all, there has been a surprising lack of discrimination in the use of terms. An eminent Washington consultant, in a recently-published article, used the terms "production per worker" and "production per man-hour" as if they were interchangeable; yet it is instantly apparent that variations in production per worker are the joint result of changes in production per man-hour and of changes in hours worked. The variation in the length of the work week over the past fifteen or twenty years, or even in the last 6 years or 6 months, has been so pronounced that this particular oversight seems inexcusable and scarcely conducive to clear thought or dependable conclusions.

Perhaps an even more outstanding characteristic of the discussions of productivity has been the widespread acceptance of a uniform rate of gain in productivity somewhere in the neighborhood of 2 per cent per year. A recent tabulation in *Dun's Review* showed the estimates of eight writers ranging from 2.0 per cent to 2.6 per cent per year increase in output per worker, with six of the eight ranging between 2.1 and 2.4 per cent. These estimates apparently are based primarily on the rather scattered and dubious data for the period from 1899 through 1919, for they ignore the fact that Dr. Fabricant's figures on output per man-hour in manufacturing from 1919 to 1939 show no evidence whatever of an exponential trend; they appear convincingly linear. Production per worker, as opposed to production per man-hour, should show even less of an exponential tendency because of the reduced length of the typical work-week over the past twenty years.

This common assumption of a persistent rate of gain in productivity opens two questions: first, the acceptability of long-range extrapolations on an exponential basis of a trend which for the latest twenty years has appeared to be arithmetic, and second, whether there is really any inherent and independent trend toward increased output per man-hour or per worker, or whether this increase may be a resultant of independent forces, lacking which the gains in productivity might cease. It seems to me more reasonable to consider increased efficiency as a result than as an inherent attribute of workers.

In this connection we might note the implicit assumption in much of the discussion of productivity that production per man-hour is primarily a function of the mental and emotional attitude of the worker toward his job.

This belief that productivity is an attribute of the individual probably stems from two sources. First, it is a commonplace of factory experience that employee "slowdowns" can seriously hinder production, while high morale and general enthusiasm can appreciably increase it.

These facts, of course, apply particularly to short intervals of time rather than to periods several years apart; they represent fluctuations about a general trend, not the trend itself.

Second, in activities where skill and craftsmanship play an important part, we have all observed the vast differences between experts and beginners. A skilled painter can paint my whole house about as fast as I can do one room—and do a better job of it, too. A skilled carpenter can make my best efforts look clumsy and futile. A good assistant can cut my work in half, while a poor one only adds to my woes. But note that all of these pursuits are non-mechanical. A good painter with a spray-gun can far outperform a good painter with a brush; a good carpenter with power tools can outperform a good carpenter with hand tools, and a good tabulating machine operator can far outdo the most diligent longhand tabulator. The more skill and effort that can be transferred from the worker to his tools, the less imposing does individual artistry and attitude become. In this age of mechanization the trend has been increasingly away from the skill and volition of the individual worker toward the speed, power and consistent performance of the machine. This trend has been especially pronounced in manufacturing.

In exploring the possible explanations of increased production per man-hour in manufacturing, we find numerous possibilities. On the side of mechanization are the application of increased power per worker and improvements in machines—greater speed, precision and specialization, with expanded use of jigs and fixtures to transfer the requisite skill and consistency of performance as far as possible from the worker to the machine. Under the heading of industrial techniques are improved designs, simplification and standardization of products, refinements of materials, development of new processes (for example, extrusion, die-casting, powder metallurgy, electronic heat-treating, etc.), better plant layout and flow of work, improved material-handling equipment, job analysis, time and motion study, quality control and many others. Affecting the performance of the individual employee are improved lighting and ventilation, higher standards of health and safety, job training for the development of skills, sundry security and recreation programs to improve morale and incentive programs to induce maximum application and effort.

Out of this list it is immediately apparent that the skill and application to duty of individual workers must play a relatively small part in the total changes in manufacturing output per man-hour over an extended period. In fact, while Dr. Fabricant's figures indicate that manufacturing output per man-hour more than doubled between 1919

and 1939, it is difficult to believe that the average skill and application of several million manufacturing workers could have undergone any such proportionate change in those two decades.

It does appear, on the other hand, that the much greater use of power per wage earner in manufacturing accounts for a very large

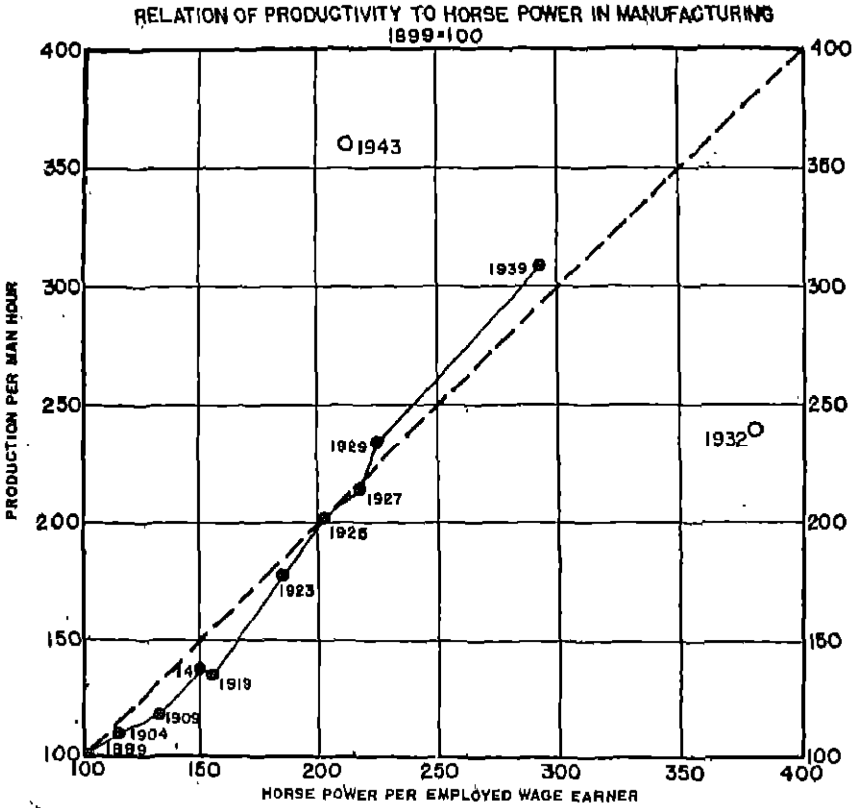


FIGURE 1

proportion of the increased output per man-hour. Taking 1899 as 100 in each case, Fabricant's figure for output per man-hour in manufacturing was 309 in 1939, while Census data on installed horsepower and wage earners in manufacturing showed horsepower per wage earner at 202. Using only the terminal years of this forty-year period, installed horsepower per wage earner in manufacturing increased 95 per cent as much as did output per man-hour.

The figures for these and other years are shown in Figure 1. The

dashed line indicates a point-for-point relationship; the sharper slope of the data since 1919 indicates that other factors than horsepower per wage earner have also operated to raise productivity, but the general similarity is still striking. For eight different combinations of terminal years, the median ratio of the increase in horsepower per wage earner to the increase in output per man-hour is 80 per cent.

One weakness in this approach is illustrated by the fact that the data between 1932 and 1943 indicate a nearly-perfect negative correlation. The answer is that in 1932 the installed horsepower figure greatly overstated the horsepower actually in use, while in 1943 a large part of the installed horsepower was being used by successive shifts of workers. A full verification of the thesis would require data (which are not available) on horsepower *actually in use* per *regular-shift* wage earner. However, the consistency of the relationship in years which were marked by neither sharp depression nor general multiple-shift operation is indicative of at least a large degree of co-variation.

A further assumption underlying this relationship is, of course, the intelligent application of power. A five-horsepower motor in the middle of the president's office would have little value, nor would the application of a 100-horsepower motor to a two-horsepower job. In general, however, it appears that horsepower in manufacturing has been intelligently applied, and that this intelligent application of power has contributed very substantially to the increased output per man-hour in manufacturing.

As a partial verification of this thesis, installed horsepower in manufacturing times the hours it is used should provide a close approximation to manufacturing output. Here again it is necessary to resort to makeshift figures, using total installed horsepower instead of horsepower actually in use, and using weekly hours per wage earner as a substitute for weekly hours per horsepower. (For non-Census years, the installed horsepower figures were interpolated *via* correlation with Department of Commerce data on expenditures for new manufacturing equipment.) The results are shown in Figure 2.

All the data in Figure 2 are index numbers based on 1899 as 100. The solid line is the product of an index of installed horsepower in manufacturing and an index of actual weekly hours per wage earner in manufacturing. The broken line is an index of manufacturing output, as reported by Fabricant, from 1899 to 1939, and by the Federal Reserve Board since 1939. Three areas of divergence merit special attention. First, between 1909 and the early 1920's there is a persistent gap of about 20 points between the two lines. Second, in all the periods of

sharp depression—1921, the early 1930's and 1938—the index of manufacturing output drops away from the product of installed horsepower and the work week, indicating that an abnormal portion of the horsepower became idle in those depression periods. Third, in all the peak periods—1929, 1937 and the 1940's—the index of manufacturing output rises above the product of installed horsepower and the work

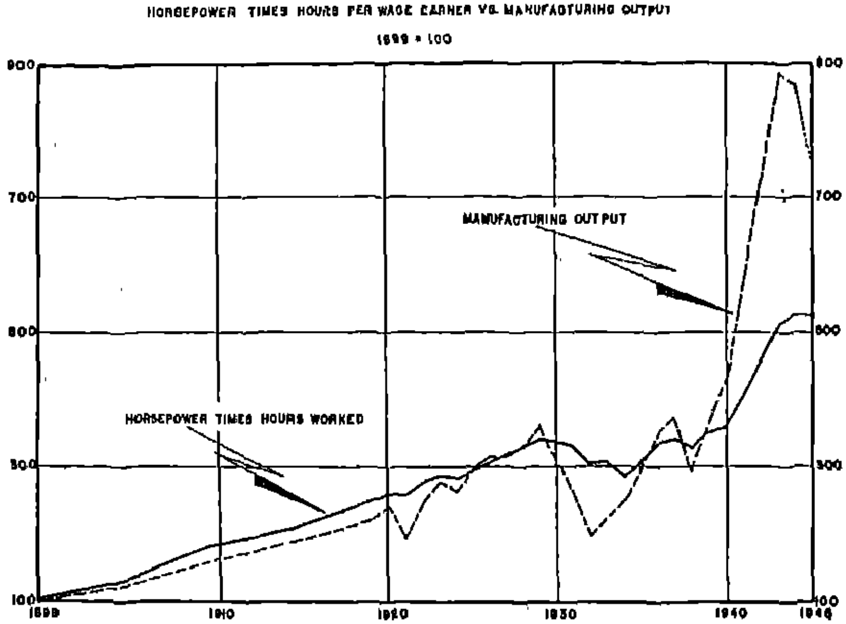


FIGURE 2

week, indicating that an abnormal portion of the horsepower was used on second and third shifts. Taking the figures at face value for the moment, they imply that during the peak years of war production one-third or more of manufacturing output was produced on extra shifts.

This conclusion results directly from the thesis that regular-shift capacity can be measured by installed horsepower and the hours per week it is used. The wartime spread between manufacturing output and the product of horsepower times hours worked per week provides a rough measure of production on shifts not ordinarily worked. The spread amounts to 371 points of the peak output index of 887 (on an 1899 base) reported by the Federal Reserve Board. This implies that

41.8 per cent of the peak output was represented either by extra-shift work or by wartime distortions of the index. If we accept the view expressed in a later paragraph that the wartime peak in the Federal Reserve Board's index of manufacturing output may have been overstated by as much as 15 per cent (a possibility that is difficult either to prove or disprove), then the portion of peak output attributable to above-normal use of equipment becomes 33 per cent instead of 42 per cent. Some portion of even this 33 per cent is undoubtedly attributable to abnormally intensive use of equipment on regular shifts, as well as to extra shifts.

Since such a large portion of wartime manufacturing output apparently resulted from extra shifts, it follows that the plant capacity of the country is not nearly adequate to support a level of activity approaching the wartime rates on the single eight-hour shift characteristic of peacetime production. Considerable extra-shift work will be involved if the full-employment levels of activity so generally discussed are attained within the next year or two, as I believe they will be.

Reverting to the relationship between horsepower per wage earner and manufacturing output per man-hour, it again appears that existing horsepower are not adequate to support even prewar levels of productivity under postwar full-employment conditions without resort to the double and triple use of horsepower involved in extra shifts.

There is little question but that a still closer relationship could be established if we were able to establish for each period the number of horsepower actually in use and the hours of their use—that is, get a measure of horsepower-hours. While in this form the project sounds imposing, it might be approached through careful interpretation of data on kilowatt-hours consumed in manufacturing, which are easily convertible into horsepower-hours.

The major problems involved would be isolating the records for power consumed in manufacturing plants and screening out the amount of power consumed in lighting and other loads not varying directly with productive activity. That power consumption is a practical measure of productive activity is illustrated by the fact that numerous plant managers keep a close check each morning on the power load as a gauge of the total activity in the plant.

It may be of interest to note, in passing, that the Federal Reserve Board's index of manufacturing output in the peak war years of 1943 and 1944 was about 15 per cent higher than a corresponding index of kilowatt-hours consumed in manufacturing based on Federal Power Commission data; the discrepancy developed after 1941, and was

considerably greater for the revised figures for 1942 than for the figures first reported.

With both indexes at 100 for 1939, manufacturing output (excluding mining) stood at 150 in 1941 and 170 in 1942, while power consumed in manufacturing was 147 in 1941 and 173 in 1942. The subsequent wartime revision of the Federal Reserve Index, however, lifted these figures to 154 in 1941 and 194 in 1942, or considerably above the power consumption index. By 1943 the Federal Reserve Index showed manufacturing output at 237 per cent of the 1939 base, or roughly 15 per cent above the power consumption index of 207.

In the search for an index of industrial activity that can be checked and verified from various angles, the available information on electric power consumed in manufacturing should be brought more prominently into the picture.

The question of productivity is closely related to some of the most burning economic issues of the day. A concerted drive is on to increase the purchasing power of organized labor, and the accent on *real* purchasing power, in terms of goods and services, is evident in the insistence that wage increases shall not be reflected in price advances. Yet from the past record the evidence is clear that advances in real hourly wages have been closely tied, for at least three decades, to advances in output per man-hour. The data are shown in Figure 3. Here again the dashed line indicates a point-for-point relationship. The data are index numbers on a 1939 base; the real income data are Bureau of Labor Statistics figures on average hourly earnings in all manufacturing divided by the BLS cost-of-living index, while the output per man-hour figures are the same as were used before, from Dr. Fabricant, extended by Federal Reserve Board and BLS data after 1939.

In general, it is apparent that there must be some fairly close relationship between the quantity of goods and services which an average hour's work can buy and the quantity which an average hour's work produces.

Since the relationship among wages, prices and productivity is so close in this combination, it should not be surprising to find it close in another. Figure 4 tells the story of the relationship between unit labor cost and finished goods prices. The solid line shows the annual average of unit labor cost for all manufacturing, while the broken line shows the annual averages of the Bureau of Labor Statistics index of finished goods prices. The unit labor cost figures were derived by dividing

average hourly earnings in all manufacturing by the index of output per man-hour in all manufacturing. Both lines are on a 1939 base.

These two illustrations point to the futility of attempting to force, solely through increases in money incomes, increases in *real* income that are not firmly rooted in high productivity. Such efforts constitute a form of warfare between economic groups, since they can only succeed by depriving one group to enrich another. Further, taken in conjunction with the relationship between power per worker and increased productivity, they tend to reduce further gains in productivity by discouraging further investment in power equipment. It is investment in plant and equipment that underlies both rising real incomes

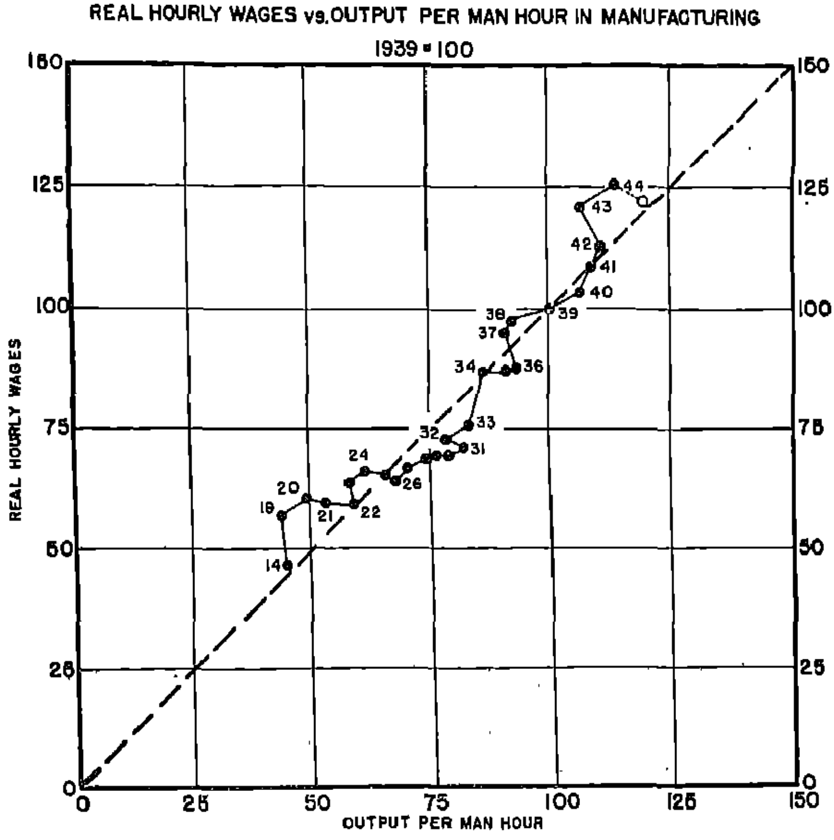


FIGURE 3

and the rising productivity whose steady annual increase has been so glibly assumed.

In view of the basic importance of productivity in appraising current and prospective economic trends, it is essential that more adequate information—and more general understanding of its significance—be developed. On the information side, the problem is primarily one of establishing satisfactory units of measurement for production and comparing the production measurements with the man-hour figures

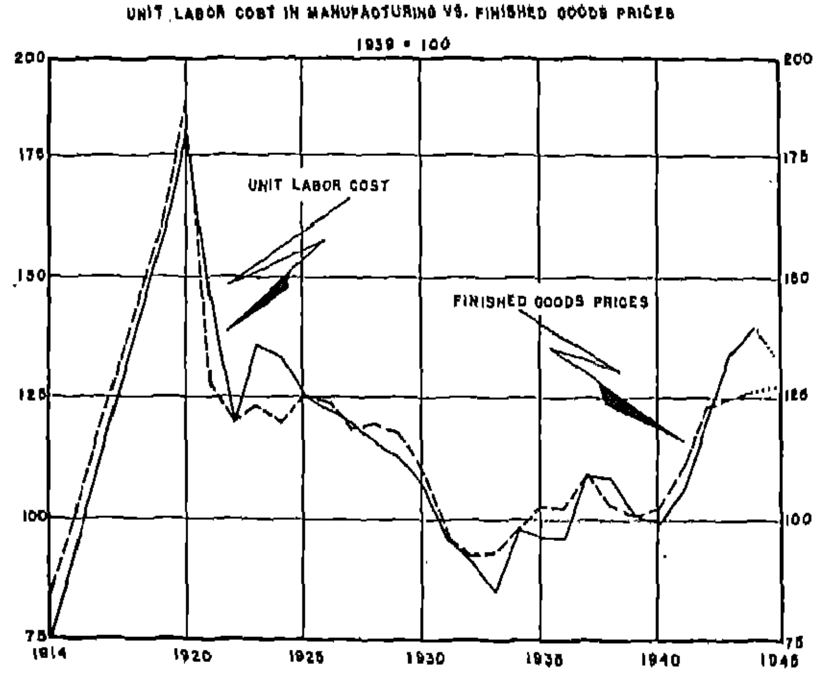


FIGURE 4

related directly to them. It is obvious that for this purpose man-hours are not satisfactory units to measure production; man-hours divided by themselves cannot trace changes in productivity. To establish production units in heterogeneous industries such as electrical equipment or chemicals it may be necessary to go into considerable detail or resort to sampling from among hundreds or thousands of individual products. The measurement of productivity, however, can be no better than the measurement of production on which it is based. Let me stress

again the desirability of greater attention to power use as a check on the reasonableness of production data.

As for widespread understanding of productivity—its origins, its significance and its limitations—any such achievement must perforce be gradual. Certainly there is much for professional economists and statisticians to learn before their findings can be broadly distributed throughout management, government and the general public. Many of the findings will run directly counter to ideas now popularly accepted, but it is a job that must be done if our democracy and our economic machine are to run on the fuel of facts—which, I take it, is why most of us are here today.

INSPECTION EFFICIENCY AND SAMPLING INSPECTION PLANS

Marvin Lavin
Corning Glass Works

The published sampling inspection plans contain the assumption that the inspection operation is completely efficient, that is, the items examined are invariably classified correctly. Some contributing factors to lack of inspection efficiency are noted, and an analysis of the validity of the guarantees of the plans in the presence of inspection error is made.

The work of Dodge and Romig, Wald, Wolfowitz, Bartky, Freeman, and other mathematical statisticians in the construction and analysis of sampling inspection plans has provided industrial quality control personnel with useful methods for attacking quality and inspection problems.

The various available sampling inspection plans provide features such as maintenance of chosen risks for wrong decisions on lot quality, minimization of inspection or of sampling, and maintenance of a selected AOQL, if the hypotheses upon which the inspection procedures are based are not severely violated in practice. To obtain the desired features of an inspection plan, the user must meet such representative conditions as randomness of sampling and the replacing of defective units, when found, by non-defectives; he, in some instances, must have *a priori* information as to the process average quality and its state of statistical control; he must provide some basis, economic or otherwise, for the choosing of tolerable risks of wrong decisions as to lot quality.

The published form of all sampling inspection plans with which the author is familiar requires the assumption that the inspection operation is completely efficient, that is, the items examined are invariably classified correctly. This requirement, it appears, is a very critical one, particularly in industries in which "visual" inspection is of prime importance. In the present state of many inspection methods, this condition is far from being met, and it is often apparent that the efficiency of the inspector in classifying individual items correctly is sufficiently low to nullify the theoretical features of the well-known sampling inspection procedures, when applied under the assumption of complete efficiency. Sustained efforts should certainly be made to remove the sources of inspection error; however, realistic analysis of the effectiveness of sampling inspection requires in many cases the consideration of the presence of repeated inspection mistakes.

People of long inspection experience have suggested that the following factors contribute to a lowered inspection efficiency (considering the simple case of rating an item as "good" or "bad," with special emphasis on visual inspection as distinct from a gaging inspection operation):

(i) The average process (or lot) proportion defective influences inspector efficiency. When the fraction defective is high, the risk of classifying a "good" item as "bad" seems also comparatively high, and the risk of classifying a "bad" item as "good" is relatively low. For a low process fraction defective, the situation as to the two types of errors is generally reversed.

(ii) The rate at which the inspector is required to work clearly affects the validity of the results of the inspection operation. Though an optimum rate usually exists, it is not always known. In addition, where there is a continuous flow of product, as on a conveyor belt, the velocity of flow changes with different items of manufacture and is often independent of the number of inspectors available to perform the operation.

(iii) The nature of the product, as well as the defects for which the inspection is being made, is often significant. Even though the maximum acceptable degree for a visual defect is clearly defined by engineering specifications and is illustrated by an adequate set of samples (needless to say, this is not always the case), many borderline cases arise requiring subjective decisions by the inspector. Moreover, some defects, by their physical nature, are difficult to discover.

(iv) A large number of personal factors also contribute to diminished inspection reliability, for example, extent of inspector experience and training, working environment, imperfect inspector vision, inspection fatigue, the inspector's conception of the importance of the defects he is considering, indifference and carelessness, etc.

To illustrate the effect of these factors, we may cite instances in which lots of ware which had been rejected by sampling inspection and subjected to a complete inspection, sometimes failed to pass a second sampling inspection, and occasionally after a second complete reinspection, failed to pass a third sampling inspection! This discouraging situation could be laid largely to low inspection efficiency, though the handling contributed some defects while the inspection operations were being performed.

To examine more specifically how inspection error may disrupt the anticipated features of a sampling inspection scheme, one is led to con-

sider the "effective" fraction defective, say p', which is related to the "true" fraction defective p by the equation

$$p' = p(1 - p_2) + (1 - p)p_1, \tag{1}$$

where p_1 is the probability of classifying a non-defective as defective, and the probability of classifying a defective as non-defective is p_2. In view of (1) in the enumeration of the factors contributing to inspection error, we note that p_1 and p_2 may be functions of p. In this discussion our interest will be centered on the case in which p_1 and p_2 are constant, that is, p' is a linear function of p,

$$p' = p_1 + (1 - p_1 - p_2)p, \tag{2}$$

and also the case in which p_1 and p_2 are related linear functions of p, say,

$$\begin{aligned} p_1 &= a + bp, \\ p &= c - bp, \end{aligned} \tag{3}$$

where again we obtain p' as a linear function of p,

$$p' = a + (1 - a + b - c)p. \tag{2'}$$

One would ordinarily have little confidence in fitting a more complicated relationship to an inspection operation.

Our fundamental concern will be with the transformation of the operating characteristic curve resulting from the introduction of the presence of inspection error. To take the simplest example, for a single sampling plan in which the sample size is n and the allowable number of defectives is c, we have the probability $L(p)$ of accepting a lot of proportion defective p as

$$L(p) = \sum_{m=0}^{c} \binom{n}{m} p^m q^{n-m}, \ (p = 1 - q) \tag{4}$$

for efficient inspection and the usual conditions for the use of the approximation (4). Introducing p_1 and p_2, we obtain

$$L'(p) = \sum_{m=0}^{c} \binom{n}{m} (p_1 + [1 - p_1 - p_2]p)^m (p_2 + [1 - p_1 - p_2]q)^{n-m}, \tag{5}$$

which defines the effective operating characteristic curve. Since

$$L'(p) = L(p_1 + [1 - p_1 - p_2]p), \tag{6}$$

one operating characteristic curve may be derived from the other in a simple graphic, or arithmetic, manner in view of assumptions (2) and (2').

As an example of the magnitude of the effect of errors in inspection on the operating characteristics of a sampling plan, we consider one based on a single sample procedure, designed to discharge the following requirements:

$$\begin{aligned} \alpha &= .1, \quad \text{for} \quad p_{t_1} = .02, \\ \beta &= .1, \quad \text{for} \quad p_{t_2} = .05, \end{aligned} \tag{7}$$

where α is as usual defined to be the maximum risk of rejecting lots of quality p_{t_1} or better, and β is the maximum risk of accepting lots of quality p_{t_2} or worse. Taking the lot size to be 10,000 (say), one finds for a sample of 235, with seven as the number of allowable defectives, these requirements are closely discharged. Table I compares the operating characteristics of this sampling plan with those of the same plan when constant inspection errors of the magnitude

$$p_1 = .02 \qquad p_2 = .05 \tag{8}$$

are assumed.

TABLE I

OPERATING CHARACTERISTICS OF SINGLE SAMPLE PLAN, $n=235$, $c=7$, WITH NO INSPECTION ERROR, AND WITH PROBABILITY OF ERROR, $p_1=.02$, $p_2=.05$

p True Proportion Defective	L_p Approximate Probability of Accepting Lot	p' Effective Proportion Defective	L_p' Approximate Effective Probability of Accepting Lot
.000	1.000	.020	.99
.005	.999+	.025	.76
.010	.997+	.029	.63
.015	.97	.034	.45
.020	.90	.039	.30
.025	.76	.043	.21
.030	.60	.048	.13
.035	.42	.053	.07
.040	.27	.057	.04
.045	.17	.062	.02
.050	.10	.066	.015
.060	.03	.076	.003
.070	.008	.085	.001
.080	.001+	.094	.000+

One direct and simple method of restoring the desired risks at the "acceptable" and "non-acceptable" quality levels immediately suggests itself. If we choose a sampling plan which maintains the chosen risks α and β for the effective fraction defective corresponding to the true acceptable and non-acceptable quality levels, we obtain the desired

operating characteristics for at least the two salient quality levels. Since

$$p_{t_2} - p_{t_1} > p_{t_2}' - p_{t_1}' \tag{9}$$

a more discriminatory sampling scheme will be required; thus one of the penalties of the presence of inspection error is the need for increased sampling.

To continue the previous example, the quality requirements (7) are transformed to

$$\begin{aligned} \alpha &= .1 \quad \text{for} \quad p_{t_1}' = .039, \\ \beta &= .1, \quad \text{for} \quad p_{t_2}' = .066. \end{aligned} \tag{7'}$$

Table II presents the operating characteristics of the single sample scheme for which the sample size is 460 and the allowable number of defectives is 23. This provides approximately the proper risks (7′), with the constant inspection inefficiency $p_1 = .02$, $p_2 = .05$. These operating characteristics should be compared with those in the second column of Table I.

TABLE II
OPERATING CHARACTERISTICS OF THE SINGLE SAMPLE PLAN,
$n = 460$, $c = 23$, WITH $p_1 = .02$, $p_2 = .05$

p True Proportion Defective	p' Effective Proportion Defective	L_p' Effective Probability of Accepting Lot
.000	.020	.999+
.005	.025	.999
.010	.029	.995
.015	.034	.97
.020	.039	.90
.025	.043	.80
.030	.048	.61
.035	.053	.44
.040	.057	.32
.045	.062	.18
.050	.066	.11
.060	.076	.02
.070	.085	.003
.080	.094	.000+

We have briefly noted the possibility of designing a sampling scheme whose operating characteristics in the environment of an estimated inspection inefficiency satisfactorily reproduces those which would be anticipated under a plan allowing for no mistakes. We may now inquire as to the channel in which corrective energies could be most profitably directed, if the existing inspection inefficiency is not to be supinely

tolerated. It is easily recognized that the two types of inspection errors cause opposing changes in the operating characteristics of a sampling plan; for some value of p, say p^*, they will exactly compensate, and this value is obtained from the relation

$$p = p_1 + (1 - p_1 - p_2)p, \tag{10}$$

that is,

$$p^* = \frac{p_1}{p_1 + p_2}. \tag{11}$$

For a chosen sampling plan, when $p < p^*$, we will have $L_p' < L_p$, that is, the probability of accepting the lot will be reduced as a result of inspection inefficiency; the converse will be true for $p > p^*$. Moreover, as p becomes farther removed from p^*, the magnitude of the difference $L_p' - L_p$ will increase monotonically. For small values of p, by writing equation (10) in the form

$$p_1(1 - p) = pp_2, \tag{12}$$

we deduce the approximate relation

$$p^* \doteq \frac{p_1}{p_2}, \tag{13}$$

and this suggests how seriously a large ratio of the two probabilities of error will affect the operating characteristics.

More generally, an examination of the equation

$$p' = p_1 + (1 - p_1 - p_2)p \tag{14}$$

reveals for the most typical industrial situations in which p, p_1, p_2 are sensibly small

$$p' \doteq p + p_1. \tag{15}$$

The principal influence is thus exerted by that class of inspection failure in which good product is rated as bad. But it is exactly this class of error which is most readily controlled. With a procedure which provides for the inspector's rejects being confirmed by a supervisor or a special reject inspector, we may sharply diminish the likelihood of such mistakes, providing the policy of additional inspection can be justified on the basis of product economics or customer relations. It may be asserted that this scrutiny will result in a proneness of the inspector to fall into mistakes of the other type, passing bad product as good, to avoid criticism. However, with $p_1 \doteq 0$, we have

$$p' \doteq p - pp_2, \tag{16}$$

indicating, by comparison with (15), that the reduction in p_1 carries more weight in restoring the expected operating characteristics than the resulting increase in p_2 further disrupts them.

No general discussion will be attempted as to the change in average outgoing quality limit which will result from inspection error. The analysis is made cumbersome under the assumption that the probability of error p_2 applies to the replacing of discarded items in samples and lots. Moreover, the $AOQL$ may be attained for a value of p which is unrealistically large. We may note, however, one simple result which arises under the hypothesis that $p_1=0$, perhaps as a result of special reject checks, and that all replaced items are indeed good product. Under these conditions, the average outgoing quality has the representation

$$AOQ' = N^{-1}[npp_2 + (N - n)pL_p' + (N - n)pp_2(1 - L_p')], \qquad (17)$$

where N is the lot size and the other symbols retain their earlier definitions. This reduces to

$$AOQ' = N^{-1}(N - n)p'L_p' + \frac{p_2}{1 - p_2}p', \qquad (18)$$

whose maximum is to be compared with that of the average outgoing quality

$$AOQ = N^{-1}(N - n)pL_p \qquad (10)$$

when no inspection mistakes are made. Direct examination of (18) indicates in the industrial range of p' and p_2 that the extreme value of (18) is in a few instances increased to the extent of 1% over (10), but in most cases the increase is substantially less than 1%. For example, with our earlier sampling plan, $n=235$, $c=7$, for which the $AOQL$ is approximately 1.9% under non-erring inspection, we obtain an increase on the order of ½%, when p_2 is assumed as large as .1 (providing the lot proportion defective remains in a limited range, say less than .20).

Viewed broadly, then, the results of our discussion on the effects of inspection failures upon sampling inspection plans suggest that it is largely sufficient to eliminate the rating of non-defective product as defective in order to have assurance that the guarantees of the plan are substantially obtained.

ELASTICITY OF PHYSICAL QUANTITIES AND FLEXIBILITY OF UNIT PRICES IN THE DIMENSION OF TIME*

FREDERICK C. MILLS
National Bureau of Economic Research

When the interaction of the physical quantities and the unit prices of a given commodity is marked by regularities over time, whether in the recurrence of cyclical or seasonal patterns or in the persistence of secular relations, it is convenient to define the differential movement of quantity with respect to price by a coefficient of elasticity, of price with respect to quantity by a coefficient of flexibility. These are mathematically identical with the familiar coefficients of Marshall and Moore, but in this usage they relate to designated temporal frameworks. The examples presented in the present paper deal with price-quantity interrelations in business cycles, in specific cycles in physical quantities, in specific cycles in unit prices, and in homogeneous secular periods. No attempt is made to eliminate or hold constant the dynamic influences that play upon the market. In each case we seek to measure the differential responsiveness of the quantity and price factors to the various impinging forces that operate within the temporal framework in question. If there are persistent regularities in the operation of these forces, appropriately designated coefficients of elasticity and flexibility can serve as useful instruments in the analysis of economic change.

I

Time and the concept of elasticity. The concept of elasticity, first formulated by Cournot in defining the law of demand, was given its familiar form by Alfred Marshall. In conventional usage the term is applied generally to the responsiveness of the demand for a given commodity to changes in the price of that commodity. More precisely, elasticity of demand is defined as the ratio of the relative change in quantity demanded (of a given commodity) to the corresponding relative change in price, when the relative changes are infinitesimal. The demand function and a derived coefficient of elasticity are assumed to relate to a moment of time, or to a period within which there are no changes in tastes, technology, income distribution, or other factors that might modify the functional relationship between quantity demanded and unit price.

* I am indebted to Maude R. Pech for aid in the preparation of this paper, and to H. Irving Forman for the construction of the charts.

The conceptual importance of this instrument is very great. Only modest progress has been made, however, in the empirical determination of demand functions and demand elasticities. Observations on commodity prices and quantities purchased are ordered in time, and thus inevitably reflect the play of dynamic factors that tend to modify the shape and location of demand functions. Great ingenuity has been shown in the attempt to measure the influence of these dynamic factors and to eliminate their effects on the recorded observations. Useful approximations to the demand curves of theory have, indeed, been achieved for a limited number of commodities, but uncertainty attaches even to these, and generalization of the results is hazardous. These efforts will continue and further progress will be made. It is probably inevitable, however, that residual traces of dynamic factors will remain in the best of these efforts and that major temporal modifications of attendant conditions will complicate empirical findings on demand functions for most commodities.

The present paper deals with investigations looking in another direction. In the study of economic processes, whether our concern be with cyclical movements or with long-term changes, the interrelated movements of unit prices and physical quantities are of central importance. These are movements in the dimension of time. They are affected by the diversity of forces acting upon both variables, as well as by the direct cause-and-effect relations between these variables. We can do something, in the choice of the frame of reference, to organize and standardize these forces, but they cannot be held constant while the specific and presumably static relations between price and quantity are studied. The diverse impinging forces, as they operate over time, are part of the record available to us. The relations between price and quantity that we may study are not timeless, then, nor are they isolated interactions. To the extent, however, that there are consistent patterns in these interactions, patterns that repeat themselves with some degree of regularity, they are of clear concern to the economist.

In this investigation we shall make use of a measure of elasticity mathematically identical with Marshall's coefficient of elasticity of demand. We shall measure the elasticity of quantity, with reference to unit price, in various temporal frameworks. The phrase "with reference to unit price" will not mean that changes in quantity are a function of changes in price, in any causal sense. In some cases, perhaps in most cases, the quantities and prices of given commodities will be responding concurrently to the pressure of outside forces. We may say that here the measure of elasticity of physical quantities defines the *differential responsiveness* of quantity to these forces. It is, as in traditional usage,

a ratio of a relative change in quantity to the corresponding relative change in price, but the quantity change will not necessarily be attributable to the price change. In some cases it will be in part attributable to the change in unit price and in part to the play of other forces, but it will not be possible to distinguish these two elements. Thus the related changes in quantity and price may not be thought of as reflecting movements along fixed demand curves (or fixed supply curves).

There would be little justification for this usage, and little hope of deriving significant results, if the time periods over which related quantity and price changes are to be studied were selected haphazardly. Under such circumstances coefficients of elasticity would be expected to change erratically, as different time units and different periods of time were employed. The justification must be found in the choice of time periods within which there is some regularity in the action of the impinging forces and in the related movements of commodity quantities and prices. We shall have more to say about the choice of temporal frameworks within which price-quantity and similar relations are to be studied.

The quantities to be employed in studying the relations between commodity prices and quantities over time need not be restricted to quantities demanded or consumed. Various series defining changes in physical volumes—amounts produced, delivered, exported, ordered—may be used. For this reason, and for others, the results are not to be thought of as in any way approximating traditional demand functions or coefficients. *Elasticity* is a meaningful and desirable term, for present purposes, but the elasticity measured is not that of demand. Indeed, it is most important that the term *elasticity of demand* be restricted to the conventional meaning. To avoid confusion, appropriate qualifying adjectives will be employed in the present paper to indicate the type of elasticity involved and the temporal framework employed in given instances. In addition, a symbol other than the customary η will be used to designate measures of temporal elasticity.

One further draft will be made upon conventional terminology. Henry L. Moore has used the term *flexibility of price* for the ratio of a relative change in the price of a commodity to the corresponding relative change in quantity.[1] The coefficient of price flexibility is the reciprocal of the coefficient of elasticity of demand.[2] It is derived, of course, when price is treated as the dependent variable. We shall use flexibility

[1] "Elasticity of Demand and Flexibility of Prices," *Journal of the American Statistical Association*, March 1922, pp. 8–10.

[2] This is not necessarily true when the two coefficients are empirically determined, with correlation less than unity.

of price as the reciprocal of elasticity of physical quantities, with qualifying adjectives to indicate the temporal framework to which given measurements relate. A symbol other than φ, introduced by Moore, will be employed, to avoid confusion with his coefficient.

II

Temporal periods for the study of price-quantity relations. Various instruments are available for measurement of the concurrent movements of physical volume and unit prices, as these elements change together over time, in mutual interaction or in response to outside forces. Our present concern is with two. The coefficient of elasticity is given by

$$e = \frac{dx}{dy} \cdot \frac{y}{x}$$

where x is a measure of physical quantity, y of corresponding unit prices. The coefficient of price flexibility is given by

$$f = \frac{dy}{dx} \cdot \frac{x}{y} .$$

The relationships are the conventional ones. The novel problem arise in the choice of temporal frameworks within which related changes i quantity and price are marked by regularities that give them economi significance. The following frames of reference give promise of yieldin significant patterns.

Business cycles. A reference framework defined by turning point (lows or highs) in general business activity in the United State has been set up by the National Bureau of Economic Researcl in connection with its general study of business cycles.[3] A con plete business cycle runs its course between successive lows (c highs). When a given series is corrected for seasonal variations th adjusted measurements on that series for one of these periods r

[3] These turning points, for the period since the Civil War, are set forth below. For explanation the derivation of the lows and highs, see *Measuring Business Cycles*, by Arthur F. Burns and Wesley Mitchell, National Bureau of Economic Research, 1946.

Trough	Peak	Trough	Trough	Peak	Trough
Dec. 1867	June 1869	Dec. 1870	Aug. 1904	May 1907	June 1908
Dec. 1870	Oct. 1873	Mar. 1879	June 1908	Jan. 1910	Jan. 1912
Mar. 1879	Mar. 1882	May 1885	Jan. 1912	Jan. 1913	Dec. 1914
May 1885	Mar. 1887	Apr. 1888	Dec. 1914	Aug. 1918	Apr. 1919
Apr. 1888	July 1890	May 1891	Apr. 1919	Jan. 1920	Sept. 1921
May 1891	Jan. 1893	June 1894	Sept. 1921	May 1923	July 1924
June 1894	Dec. 1895	June 1897	July 1924	Oct. 1926	Dec. 1927
June 1897	June 1899	Dec. 1900	Dec. 1927	June 1929	Mar. 1933
Dec 1900	Sept. 1902	Aug. 1904	Mar. 1933	May 1937	May 1938

flect the play of forces associated with cycles in general business, of trend factors and random influences, and, if they be present, of specific cyclical factors independent of cycles in general business. If observations cover a number of business cycles the effects of random factors may be in some degree eliminated by averaging. (Intra-cycle trend is not removed.) It is possible, then, to determine whether there is a persistent pattern in the behavior of a given series or in the related behavior of paired series within the reference framework provided by cycles in general business.[4] Within this framework we may study the differential responsiveness of the volume factor with reference to cyclical changes in unit prices (the cyclical elasticity of physical quantities) or the differential responsiveness of unit prices with reference to cyclical changes in volume (the cyclical flexibility of prices).

Specific cycles in quantities. Many physical volume series are marked by cycles specific to the particular series, cycles that may differ materially in timing and duration from cycles in business at large. Hog cycles and construction cycles are familiar examples of such distinctive rhythms; similar specific cycles are discernible in the records for many commodities. These cycles provide obvious frameworks within which the movements of prices with reference to concurrent quantity changes may be studied.

Specific cycles in prices. Price series have their own distinctive cycles, for many commodities. The elasticity of quantities with reference to price changes may be studied in these frames of reference. In this case, as for general business cycles and specific cycles in quantities, there is no isolation of price and quantity movements. Other forces than those of price play upon quantities when specific price cycles provide frames of study. This is an unavoidable condition, with the data available to us. But price factors are given primary place, in so far as this is possible, in the selection of this framework, just as quantity factors are given primary place in the use of specific cycles in quantities as frameworks for the study of price flexibility.

Long cycles. If the reality of long cycles can be established, these will provide frames similar to the reference cycles derived from the records of business cycles. The elasticity of quantities and the flexibility of prices within these cycles could then be measured, and their significance tested.

Homogeneous secular periods. In defining each of the cyclical periods

[4] Measurements of the behavior of a given series during a business cycle are said to define a *reference cycle*.

named above we have sought to set off a section of time marked by some regularity in the forces at play, particularly the forces affecting the two variables with whose relative changes we are concerned. Regularity may be tested, for cyclical measures of this sort, with reference to the degree of variation found among observations for different cycles. Such tests are not possible when we deal with periods of nonrecurring change, but such periods are of central interest to the student of economic change. There are stages of economic development marked by the play of fairly persistent secular forces, these stages giving way to others during which quite different secular influences are operative. The years 1914–19 marked such a secular turning point in many aspects of economic life in the United States. We may not set down precise criteria of secular homogeneity, but certain stages of development may be defined with reasonable accuracy. Interactions of quantities and prices, and their differential responsiveness to secular forces, may be measured for such periods by coefficients of secular or intercycle elasticity and flexibility. The significance of such measures is not open to exact statistical determination, but comparisons of measures for different periods and consideration of these measures in their full economic setting for given periods will provide bases for judgments as to their economic import and value.

Special periods subject to the play of defined forces. The differential responsiveness of quantities and prices to particular forces may be studied for specified periods, known to be distinctive in the forces at work. Thus the period of NRA regulation is a sharply defined section of time during which special forces were acting upon the structure of production and the system of prices. Here, although regularity of pattern in the sense of repetitiveness is not to be expected, the condition that distinctive and identifiable forces are operating is realized. For such a period measures of elasticity and flexibility of the type here discussed may be highly serviceable.

This list of possible reference frames for use in studying the interactions of quantities and prices and their responsiveness to outside forces is not exhaustive. For certain commodities markedly subject to seasonal influences (such as eggs or butter), the year would be a suitable period for study. The essential feature is that particular forces be distinguished, by the judicious selection of temporal frameworks. Persistence of operation of these forces within a given period, in the case of nonrecurring actions, or persistence of pattern in different periods, in the

case of cyclical or seasonal forces, is requisite, of course, if the measurements are to have meaning. If purely random factors predominate no such persistence is to be expected.

III

Elasticity and flexibility in business cycles. For purposes of analysis the seasonally corrected price and quantity records for a given commodity, during a given business cycle, are expressed in relative terms. The base of the relatives for a given economic series in a given business cycle is the average of the original monthly (or other) figures for that series in that cycle. Averages of these relatives are then struck for each of nine stages of that cycle. Stage I is the initial trough, stage V the peak, and stage IX the terminal trough. Stages II, III, and IV mark off successive thirds of the phase of expansion, and stages VI, VII, and VIII mark off successive thirds of the phase of contraction.[5]

Stage averages for the price of steel billets and the production of steel ingots for ten business cycles in the United States between 1900 and 1938 are given in Table 1. They are plotted by reference cycle stages in Figure 1. This chart shows the rapid advance in ingot production and the much more moderate rise in the prices of steel billets in the first stage of reference expansion, the continuing sharp increase in production and the accelerating but less rapid advance in prices during the remaining stages of expansion. The movements during contraction are generally symmetrical with those of expansion, with quantity produced declining much more rapidly than prices except in the terminal stage of the phase of contraction.

These same materials are plotted in a coordinate framework in Figure 2. Here the movements of quantities are shown on the horizontal axis, the movements of prices on the vertical axis. This representation reveals more clearly the apparent responsiveness of the quantity factor to price changes, and the apparent flexibility of prices as quantities change, within the framework of business cycles. It is clear that no assumption of direct causality is justified in the study of this relationship. We may define the relative change in steel production with reference to the corresponding relative change in steel prices between, say, reference cycle stages II and III. What we are measuring in so doing is not a change in quantity resulting from a corresponding change in price, but a change in quantity attributable to the influence of a combination of forces of which the price factor is one. Conversely, we may measure the differential responsiveness of unit prices with

[5] This procedure is described in detail in *Measuring Business Cycles, op. cit.*

reference to corresponding changes in physical quantities as both variables change in the course of business cycles. In each case the change in the "independent variable" is to be thought of as a standard of reference rather than a primary causal influence.

With these qualifications in mind it is convenient to define the relative responsiveness of physical quantities and of unit prices by means of the coefficients of elasticity and flexibility previously described. These are given in Table 2 for the several cyclical stages and phases, and for the full cycle.[6] It will be understood that these measurements are based upon the average movements of steel ingot production and

TABLE 1

AVERAGES DEFINING THE REFERENCE CYCLE PATTERNS OF STEEL INGOT PRODUCTION AND THE PRICES OF STEEL BILLETS IN TEN BUSINESS CYCLES, 1900-38*

	Stage Averages (Reference cycle relatives)								
	I	II	III	IV	V	VI	VII	VIII	IX
Quantity	70	91	106	118	128	122	98	76	74
Prices	80	87	98	100	114	113	103	94	90

* Billets and ingots are produced at different stages of the manufacturing process, but for present purposes fluctuations in the output of ingots may be taken to define with acceptable accuracy volume changes related to cyclical movements in the prices of steel billets. Ingot production is the total for the United States. Prices are for Bessemer steel billets at Pittsburgh.

steel billet prices over the ten reference cycles covered by the present observations.

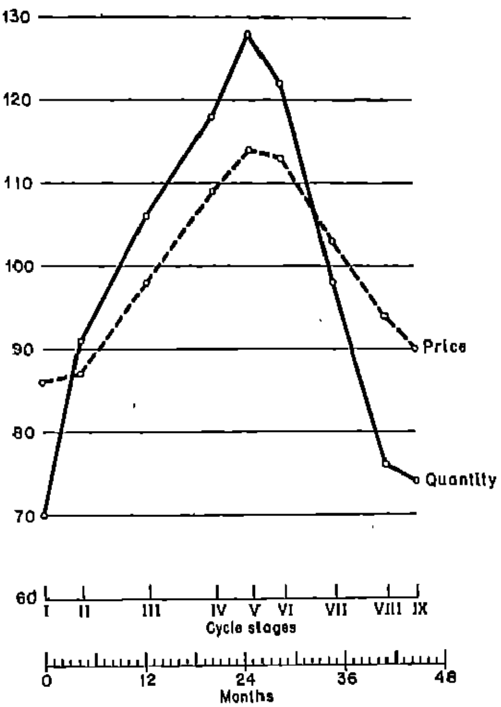
The coefficient of elasticity of steel ingots indicates that over the full reference cycle there is, on the average, a change of 2.17 per cent

[6] The coefficients are derived from the entries in Table 1. *Stage coefficients are computed from the reference cycle relatives for successive stages.* For interstage period II–III, for example, the coefficient of elasticity is *the ratio of the relative rates of change—quantity to price—at the midpoint of the line joining the quantity and price observations for stages II and III.* Thus, for this period, $e = \frac{+15}{+11} \cdot \frac{92.5}{98.5} = +1.28$. The coefficients for the expansion phase are derived from reference cycle relatives for stages I and V, those for the contraction phase from similar relatives for stages V and IX. (*A stage or phase coefficient derived from the customary formula and relating to the midpoint of a straight line joining two pairs of observations on prices and quantities is, of course, identical with a coefficient of arc elasticity.*) *Full cycle measures are averages of the coefficients for the phases of expansion and contraction.* When these phase coefficients are of the same sign, geometric means of arithmetic and harmonic averages of the two phase coefficients will give mutually consistent sets of elasticity and flexibility coefficients for the phases and the full cycle. When the phase measures are of opposite sign mutually consistent coefficients (i.e., measures of elasticity and flexibility that are reciprocals of one another for both phases and for the full cycle) may not be obtained by an averaging process. In such cases, and wherever the coefficients for the phases of expansion and contraction differ materially, it is well to place chief reliance, for interpretative purposes, on the coefficients for the separate phases rather than upon an average for the full cycle.

in production for every change of 1 per cent in price, these changes being in the same direction. The separate coefficients for the phase of expansion and the phase of contraction are but slightly different. In both phases the relative changes in output are more than double the corresponding relative changes in unit price.

The measures for the separate interstage periods are all positive,

FIGURE 1
Average Movements of Steel Ingot Production and Prices of Steel Billets in Ten Business Cycles 1900 – 1938



indicating that the concurrent changes of production and price are in the same direction in all cyclical periods. There are some notable differences in magnitude, however. In the first period of business recovery (between reference cycle stages I and II) the coefficient of elasticity has a value of +22.57. Quantities respond immediately and strongly to the first pressures of general business recovery. The rise in unit price

is slight at this stage. Similarly, the first impact of recession between reference cycle stages V and VI brings a much sharper drop in steel production than in steel prices (the coefficient of elasticity is +5.45). It is to be noted that the coefficients of elasticity for the two periods of contraction following stage VI are much greater than the coefficient of elasticity for the corresponding periods of business expansion. When contraction is well under way, quantity is relatively more responsive and billet prices are relatively less responsive than they are in corre-

TABLE 2

COEFFICIENTS OF ELASTICITY OF STEEL INGOT PRODUCTION AND FLEXIBILITY OF STEEL BILLET PRICES IN TEN BUSINESS CYCLES, 1900-38

	Stage Measures Interstage Period							
	I–II	II–III	III–IV	IV–V	V–VI	VI–VII	VII–VIII	VIII–IX
Elasticity	+22.57	+1.28	+1.01	+1.81	+5.45	+2.30	+2.76	+0.01
Flexibility	+0.04	+0.78	+0.00	+0.55	+0.18	+0.42	+0.36	+1.64

	Phase Measures Expansion	Contraction	Full Cycle Measure
Elasticity	+2.00	+2.27	+2.17
Flexibility	+0.48	+0.44	+0.46

sponding periods of business expansion. In all interstage periods except the terminal stage of recession the coefficients of elasticity exceed unity and the coefficients of price *flexibility* fall below unity. Only between reference cycle stages VIII and IX is the elasticity of output less than unity. The influence of recovery is reflected thus early in the retarded decline of physical output. Prices, however, continue to drop at a rate exceeding that prevailing during the first stage of contraction. Except for this one period, therefore, steel ingot production is elastic, positively (positively because the movements of production and price are in the same direction), and the prices of steel billets are inflexible, positively.

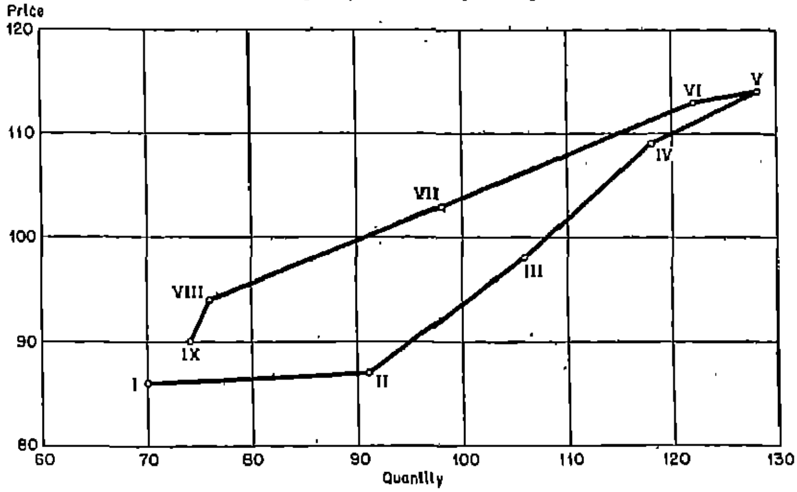
The patterns of related price-quantity behavior represented in Figures 1 and 2 are quite consistent over the period covered by the present records. Those for nine of the ten cycles are of the type represented by Figure 2, marked by price and quantity increases during general business expansions, price and quantity declines during general business contractions. Only for one of the ten cycles—that of 1924–27—is the pattern different. In this cycle we have prices declining during the phase of reference expansion instead of increasing, as in the nine other

cycles covered. Measures of conformity and variance tests indicate that the pattern of average behavior represented in Figure 2 is clearly significant.

A quite different pattern of cyclical behavior is represented by measurements of the price of raw cotton and the quantity of cotton exports (Table 3 and Figure 3). These again relate to stages of cycles in business at large. The general relationship is clearly inverse. Unit prices con-

FIGURE 2
Pattern of Steel Ingot Production and Prices of Steel Billets in Ten Business Cycles, 1900 – 1938

Averages by reference cycle stages



form to the cycle, rising during the phase of general business expansion between reference cycle stages I and V (prices rise, in fact, with a slight lag, starting after stage II), and falling during business contraction (between reference cycle stages V and IX). These are, of course, prices in the United States; they reflect the cyclical pressures present in domestic markets. The pattern of quantity movements reverses the cyclical tides (except during two interstage periods), declining during business expansion and rising during business contraction. Here, it is reasonable to assume, we have something approaching the short-term market relationships of conventional theory. The volume of cotton exports is strongly influenced by business conditions abroad and by cotton prices in American markets. Foreign business cycles are not synchronous with United States cycles and, indeed, differ from country to country. In the reference frame here set up, therefore, the influence of

foreign business conditions is variable, if not erratic; the one persistent force affecting the quantity of exports is the customary stimulating influence of falling prices upon demand, the customary depressing influence of rising prices. Accordingly, with averages covering many domestic cycles, the volume of cotton exports tends to be an inverse function of prices in the United States. This is the case for the present set of observations.

TABLE 3

AVERAGES DEFINING THE REFERENCE CYCLE PATTERNS OF COTTON EXPORTS AND COTTON PRICES IN SEVENTEEN BUSINESS CYCLES, 1870-1038*

A. Stage Averages

	(Reference cycle relatives)								
	I	II	III	IV	V	VI	VII	VIII	IX
Quantity	102	108	104	05	89	05	103	90	106
Prices	07	07	101	110	113	100	101	07	93

B. Coefficients of Elasticity of Exports and Flexibility of Prices†

	Stage Measures Interstage Period							
	I–II	II–III	III–IV	IV–V	V–VI	VI–VII	VII–VIII	VIII–IX
Elasticity	−27.71	−0.03	−1.06	−2.42	−1.81	−1.06	+1.74	−2.35
Flexibility	−0.04	−1.07	−0.94	−0.41	−0.56	−0.94	+0.57	−0.43

	Phase Measures		
	Expansion	Contraction	Full Cycle Measure
Elasticity	*−0.89*	*−0.00*	*−0.00*
Flexibility	−1.12	−1.11	−1.11

* The quantity series is that for raw domestic exports. The price is that for raw middling upland cotton in New York.

† These coefficients are computed from stage averages carried to one decimal place, where the rounded averages for successive stages are equal.

Measures of elasticity and flexibility are given in Part B of Table 3. The phase measures are nearly identical. The elasticity of foreign demand for American cotton, within the framework of domestic business cycles, is measured by a coefficient of −0.00. All but one of the interstage elasticity measures are negative; six of the eight are negative and numerically in excess of unity.

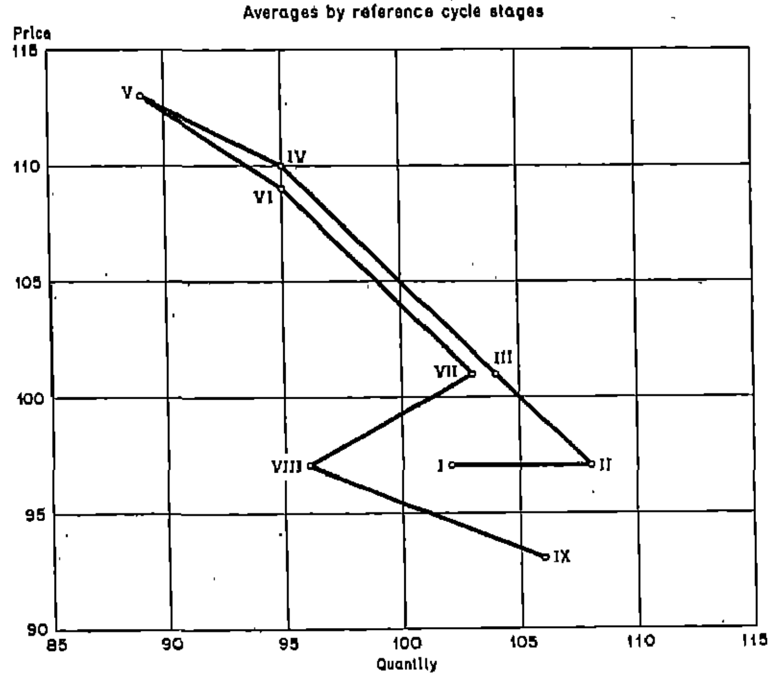
The cyclical behavior of cotton exports and raw cotton prices is distinctly less consistent than is that of steel production and unit price. The patterns for the seventeen reference cycles, viewed separately, show considerable variation. Of the seventeen, six are of the same type

as the average pattern plotted in Figure 3. They are marked by prices that move with the general cyclical tide and by exports that move inversely to that tide. The other eleven patterns differ widely in their characteristics. Tests of statistical significance indicate that raw cotton prices conform to the patterns of reference cycles in business, but

FIGURE 3

Pattern of Raw Cotton Exports and Raw Cotton Prices in Seventeen Business Cycles, 1870–1938

Averages by reference cycle stages



that cotton exports do not move in a clearly significant pattern. The most common pattern is that corresponding to the average shown in Figure 3, but there is too much variation to justify the acceptance of this pattern as clearly typical.

Measures similar to those given in Table 1 have been computed for 64 commodities.[7] Combining the observations for these individual com-

[7] These measures, and the analytical procedures employed, are described in a monograph on *Price-Quantity Interactions in Business Cycles* (National Bureau of Economic Research, 1946). The paragraphs of the text immediately following are from this study.

The sample of 64 commodities is reasonably comprehensive. It includes 32 raw materials, 32 manu-tured goods; 33 farm products, 31 nonfarm products; 48 producer goods, 22 consumer goods (including

modities we obtain measures descriptive of the average behavior of physical quantities and commodity prices at wholesale during cycles in general business. This average pattern, as defined by the entries in Table 4, Part A, is shown graphically in Figure 4. This aggregative picture, it is clear, is one of positive conformity of both prices and quantities to business cycles. In this framework general cyclical forces shape the fluctuations of both prices and quantities, overcoming any tendencies that may exist toward inverse movements.

TABLE 4

PATTERNS OF BEHAVIOR OF QUANTITIES AND PRICES DURING BUSINESS CYCLES, AVERAGES FOR 64 COMMODITIES

A. Stage Averages

	(Reference cycle relatives)								
	I	II	III	IV	V	VI	VII	VIII	IX
Quantities	89	98	101	108	112	107	99	94	94
Prices	94	99	104	110	112	110	99	91	90

B. Coefficients of Elasticity of Quantities and Flexibility of Prices

	Stage Measures, Interstage Period							
	I–II	II–III	III–IV	IV–V	V–VI	VI–VII	VII–VIII	VIII–IX
Elasticity	+1.50	+0.60	+1.34	+1.47	+2.34	+0.71	+0.99	−0.32
Flexibility	+0.63	+1.45	+0.75	+0.68	+0.43	+1.41	+1.45	−3.12

	Phase Measures: Expansion	Phase Measures: Contraction	Full Cycle Measure
Elasticity	+1.25	+0.81	+1.01
Flexibility	+0.80	+1.23	+0.99

Our immediate interest is in the relative responsiveness of the price and quantity factors to the forces of business cycles. Over the full cycle (see the last entry in Part B, Table 4) there is little difference between the two, for the sample of commodities here studied. By convention, we should classify the quantity movements as elastic (positively), the price movements as inflexible (positively), since the coefficient e is greater than unity while f is less than unity. But the difference is too small to be significant.

duplications). It is not, however, presented as representative of all commodities entering into trade.

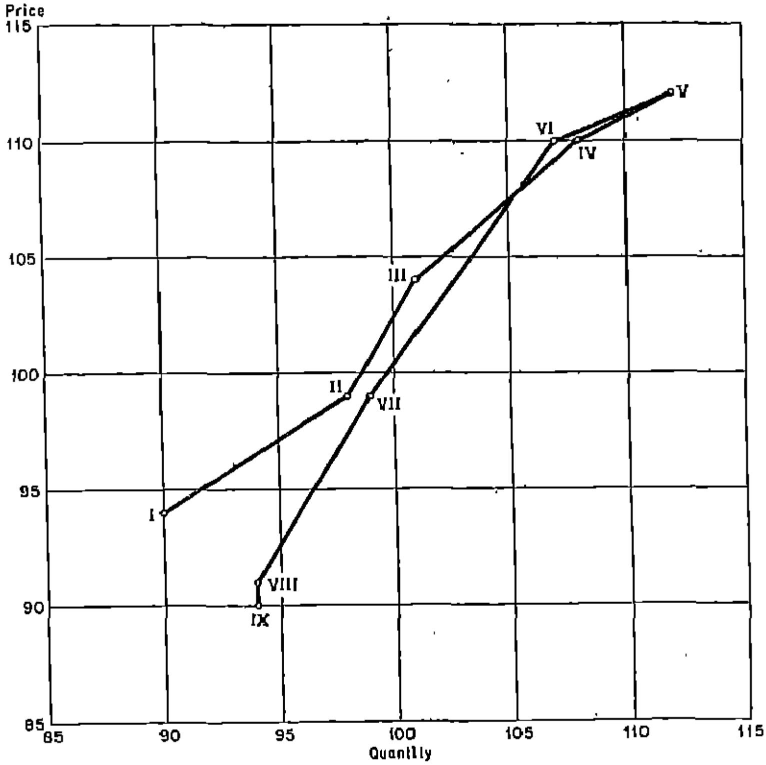
There is considerable variation in the time coverage of the series included. For two commodities we have observations going back to 1858, covering twenty business cycles; for three commodities the coverage is restricted to the three business cycles that have run their course since 1924. The other 59 commodities fall between these extremes. Observations for recent cycles are more numerous than those for earlier cycles.

If we go behind the approximate equality of the price and quantity measures for the full cycle we find notable differences in behavior from phase to phase and from stage to stage of reference cycles. The phase

FIGURE 4

Pattern of Related Price-Quantity Movements in Business Cycles
64 Commodities

Averages by reference cycle stages



coefficients tell an illuminating story of related quantity and price changes in business cycles. During periods of general business expansion physical volume increases 1.25 per cent for every 1 per cent rise in price. During contractions commodity prices decline 1.23 per cent for every 1 per cent fall in physical volume. We shall find wide differences among commodity groups in the correlated behavior of quantity and price, but for the aggregate of commodities here studied quantity is elastic, positively, in response to the stimulations of expansion, and

unit prices are flexible, positively, under the pressures of contraction.

We trace these relations more closely in the measures for interstage periods. Quantities are elastic, relatively, in three of the four interstage periods of expansion (in period II–III alone are prices more responsive than quantities to cyclical pressures), inelastic in three of the four interstage periods of contraction (in period V–VI alone are quantities more responsive than prices to cyclical pressures). All the interstage

TABLE 5

AVERAGES DEFINING THE REFERENCE CYCLE PATTERNS OF FACTORY EMPLOYMENT AND AVERAGE HOURLY EARNINGS IN FOUR BUSINESS CYCLES, 1921–38*

A. Stage Averages

	(Reference cycle relatives)								
	I	II	III	IV	V	VI	VII	VIII	IX
Employment	90	96	102	109	114	111	102	94	90
Average hourly earnings	93	95	98	101	106	107	107	103	102

B. Coefficients of Elasticity of Employment and Flexibility of Average Hourly Earnings

	Stage Measures Interstage Period							
	I–II	II–III	III–IV	IV–V	V–VI	VI–VII	VII–VIII	VIII–IX
Elasticity of employment	+3.09	+1.95	+2.20	+0.93	−2.84	+12.70	+2.14	+4.40
Flexibility of hourly earnings	+0.33	+0.51	+0.45	+1.08	−0.35	+0.02	+0.47	+0.22

	Phase Measures		
	Expansion	Contraction	Full Cycle Measure
Elasticity of employment	+1.60	+6.12	+3.32
Flexibility of hourly earnings	+0.50	+0.10	+0.30

* The employment series is based upon estimates by the United States Bureau of Labor Statistics of the number of skilled and unskilled workers in 13 to 90 manufacturing industries. Average hourly earnings are the National Industrial Conference Board estimates for skilled and unskilled workers in 25 manufacturing industries. The comparability of the two series is impaired, of course, by the difference in coverage, but the present estimates define the general cyclical pattern with reasonable accuracy.

measures are positive (indicating direct relations between price and quantity changes) except that for the terminal period of contraction (VIII–IX). At this final stage prices are declining and quantities increasing,[8] but the increase in quantity is relatively less rapid than the decline in price, and the coefficient e is below unity.

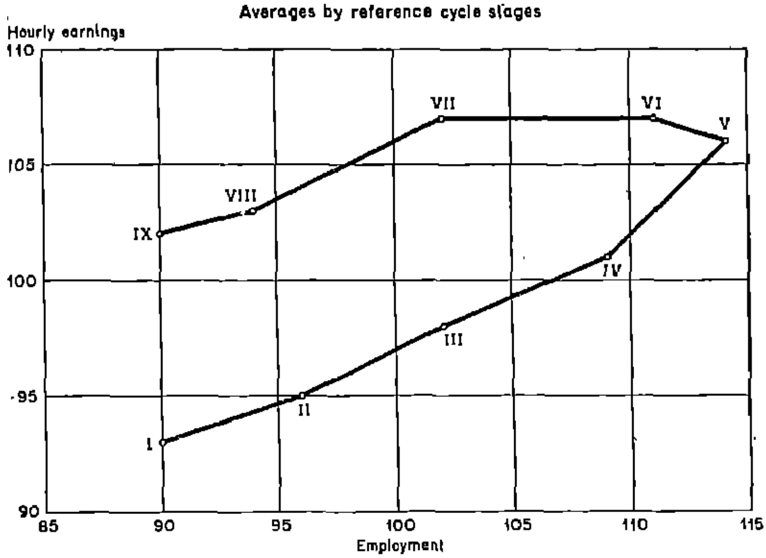
The coefficients for the eight interstage periods show a distinct and suggestive pattern of change. Following the decline in the elasticity of

[8] The measures of elasticity and flexibility in Table 4 are based upon stage averages carried to one decimal place. The slight rise in quantity between stages VIII and IX is not apparent in the rounded figures in Part A of Table 4

quantities after stage II, there is a progressive increase in the stage elasticity of physical quantities (and a corresponding progressive decline in the *flexibility* of unit prices) between periods II–III and V–VI of reference cycles. We may think of these two factors as representing alternative means by which markets respond to and adapt themselves to the pressures of rising demand during business expansion, of declining demand during contraction. After the reversal of the second period,

FIGURE 5
Pattern of Factory Employment and Hourly Earnings in Four Business Cycles, 1921–1938

Averages by reference cycle stages



accommodation is effected in increasing degree through physical output. Prices become decreasingly flexible, relative to quantities. There is a sharp contrast indeed between the situation prevailing between stages II and III, when prices advance 1.45 per cent for every 1 per cent rise in quantity, and that in the final period of reference expansion (IV–V) when prices advance only 0.68 per cent for every 1 per cent rise *in quantity*. These growing strictures on prices, relative to the forces affecting output and sales, are a notable feature of the present evidence bearing on the later stages of business expansion.

In the first period of contraction (V–VI) quantities drop sharply, while prices decline but slightly. Thereafter the record is the opposite of that for expansion. There is a pronounced and progressive decline

in the elasticity of quantities (and a progressive increase in the flexibility of unit prices). Quantity falls at a declining rate, relatively to prices; prices decline at accelerating rates, relatively to quantities. Between stages V and VI prices decline 0.43 per cent for every 1 per cent fall in quantities; between stages VII and VIII average unit prices fall 1.45 per cent for every 1 per cent fall in quantities. As general contraction spreads, and pervades the economy, resistances to continuing reductions in physical quantity are stronger, relatively, than the resistances to continuing price declines. For the sample of commodities here represented price is distinctly the more responsive factor, in all except the first period of business contraction.

We deal, finally, in the reference cycle framework, with the volume of factory employment and average hourly earnings of factory workers. For these series we have observations covering four reference cycles between 1921 and 1938.[9] Reference cycle relatives, averaged by cyclical stages, are given in Table 5 together with derived measures of the elasticity of factory employment and the flexibility of hourly earnings. The joint behavior pattern is plotted in Figure 5.

Employment is, of course, relatively more responsive to cyclical pressures than are hourly earnings. For the full cycle the elasticity of employment is defined by a coefficient of +3.32. For every change of 1 per cent in earnings there is a concurrent change, in the same direction, of 3.32 per cent in the number of workers employed. The expansion and contraction phases differ materially, however. The elasticity of employment averages +1.80 during periods of business expansion, +6.12 during contractions. In both phases the volume of employment is more responsive than earnings to the forces of business cycles, but during contraction the relative change in employment is sharply accentuated. The comparative inflexibility of average hourly earnings in contraction accounts for the difference, since the amplitude of employment movements is the same in the two phases.

The pattern of interstage changes is revealing. We note the relative elasticity of employment, the relative stability of earnings as general business expansion gets under way. The employment factor remains elastic, and hourly earnings relatively inflexible, through stage IV. The pressures of rising demand for labor are met, during most of the expansion phase, by sharp advances in the number employed, by

[9] Measurements descriptive of the joint cyclical behavior of employment and hourly earnings in these four cycles may not be generalized without additional evidence. The period covered is short, and it is known that special circumstances influenced the relations in question during this period. We may note, however, that customary statistical tests indicate that the average cyclical pattern of employment changes is significant; hourly earnings show high conformity for the phase of expansion and for the full cycle, but there is no clear evidence of significance.

small advances in hourly earnings. But in the terminal period of expansion (between stages IV and V) the flexibility of earnings exceeds unity, while the elasticity of employment is less than unity. In the first stage of contraction the coefficients are negative, reflecting the persistence of advances in hourly earnings, although volume of employment has started to drop. The resistance of earnings to downward pressures is manifest in a flexibility measure close to zero for interstage period VI–VII. Thereafter, during the two remaining periods of business contraction, the coefficients of flexibility are positive (i.e., hourly earnings decline as employment falls), but they remain well below unity. Hourly earnings are far less responsive to the forces of business contraction than to the forces of expansion.

IV

Measures of elasticity and flexibility relating to movements in the framework of business cycles have specialized meaning and uses. The relatives on which they rest do not, in most cases, reflect the full cyclical swings of the two factors, for the imposition of a standard reference frame will dampen or eliminate price or quantity fluctuations not synchronized with cycles in general business. To supplement the reference cycle record the fluctuations of commodity prices must be studied within a frame provided by specific cycles in the corresponding physical quantity series, and the fluctuations of physical quantities must be studied within a frame provided by specific cycles in prices. The three sets of records are needed to define cyclical market behavior.

Specific cycles as frames of reference: Flexibility of prices during specific cycles in quantities. Between 1865 and 1937 seventeen specific cycles in the physical volume of hog receipts at Chicago may be identified. In the present analysis of this series, and in the related study of hog prices, the reference frame is provided by the cyclical turning points (i.e., the lows and highs) in the series of hog receipts itself. Phases of expansion and contraction and cyclical stages (from I to IX) are identified with the series of hog receipts, not with cycles in business at large. For some commodities, of course, specific cycles may coincide with or closely approach general reference cycles. Specific cycles in hog receipts do not conform to reference cycles.[10]

Stage measurements for hog prices and hog receipts, within the specific cycle framework for hog receipts, are given in Table 6 and are plotted in Figure 6. These are averages for seventeen specific cycles.

[10] The index of conformity is −10, on a scale that runs from −100 to +100. It is not significantly different from 0, which would represent no conformity, direct or inverse.

Conformity and variance tests indicate that the pattern of price changes, within this framework, is significant. The general picture is of an inverse relationship, with prices falling as receipts increase and rising as receipts diminish. In fitting the price series into the quantity framework we are giving the quantity series an independent status. We are, in effect, setting up the hypothesis that price is a function of quantity.[11]

TABLE 6

AVERAGES DEFINING THE PATTERN OF HOG RECEIPTS IN SEVENTEEN SPECIFIC CYCLES AND THE PATTERN OF HOG PRICES IN THE QUANTITY FRAMEWORK, 1866-1937*

A. Stage Averages

	(Cycle relatives, quantity framework)								
	I	II	III	IV	V	VI	VII	VIII	IX
Quantity	75	87	101	111	104	112	100	92	82
Prices	118	104	91	91	91	99	111	114	119

B. Coefficients of Flexibility of Prices in Quantity Framework

	Stage Measures Interstage Period							
	I–II	II–III	III–IV	IV–V	V–VI	VI–VII	VII–VIII	VIII–IX
Flexibility	−0.85	−0.90	0	0	−0.47	−2.08	−0.18	−0.37

	Phase Measures Expansion	Contraction	Full Cycle Measure
Flexibility	−0.40	−0.55	−0.50

* The quantity series defines hog receipts at Chicago, by number. Prices are for heavy hogs at Chicago, in 100 pound units.

Stage and phase measures of the flexibility of hog prices, in response to cyclical changes in hog receipts, are given in Part B of Table 6. The coefficient for the full cycle is −0.50; the separate phase measures for

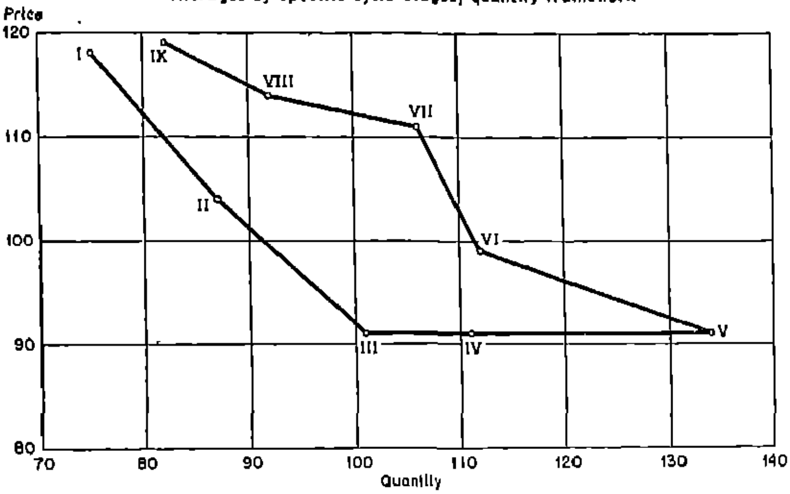
[11] This involves the assumption that, within the framework of specific cycles in quantities, causal relations run from changes in hog receipts to changes in prices. If the quantity series conformed closely to reference cycles the influence upon prices of specific variations in quantities would be indistinguishable from the influence of cyclical forces at large. But when the specific quantity cycles are apparently unrelated to reference cycles, as in the present case, the assumption that price changes within the quantity framework reflect the influence of quantity changes is more tenable. Mutual interaction of prices and quantities is not ruled out and, it is understood, the influence of other forces is never clearly eliminated. However, the process of averaging observations covering many quantity cycles gives opportunity for the offsetting and cancellation of forces not persistently operating in that specific framework.

The hypothesis that prices are independent, with quantity fluctuations playing a dependent role, can be tested by studying the movements of hog receipts within a framework provided by specific cycles in hog prices. (This is done, for another commodity, in the example next following.) Since there is significant conformity of quantities within the price framework there is no clear basis for rejecting either hypothesis. This conclusion is consistent with the assumption that there is mutual interaction of hog prices and hog receipts within the present context of short- and medium-term market relations.

expansion and contraction do not differ materially from this. In general, for every change of 1 per cent in the volume of hog receipts, during specific cycles in such receipts, there is an inverse change of about one-half of 1 per cent in prices. The interstage measurements indicate a fairly persistent pattern of price response to quantity changes, but one notable feature is to be remarked. The responsiveness of prices to increases in hog receipts is relatively strong (although the coefficients remain below unity) between interstage periods I and III. Receipts

FIGURE 6
Pattern of Hog Receipts and Hog Prices in Seventeen Specific Cycles in Hog Receipts, 1865–1937

Averages by specific cycle stages, quantity framework



continue to increase between stages III and V but prices do not decline. Whether this apparent drop in price flexibility represents accidents of sampling or a persistent economic characteristic of hog markets is a matter for investigation.[12] In the contraction phase of specific cycles in hog receipts, prices are most flexible during the first two periods (between stages V and VII). When quantities decrease, as when they increase, hog prices tend to move inversely, the tendency being strongest during the first stages of decline and advance. Thereafter, prices become less responsive to continued movements of quantities.

Elasticity of quantities during specific cycles in prices. Specific cycles in

[12] Records of price flexibility in response to changes in hog slaughter, covering 14 specific cycles in the slaughter records, reveal a similar sharp decline in flexibility after stage III of the expansion phase.

prices may be used as a framework within which the behavior of physical volume series may be studied, just as quantity cycles provide a framework for the study of prices. As an example we study the pattern of lead ore shipments within a frame provided by specific cycles in the prices of lead ore. Measurements derived by averaging data for eleven specific cycles in lead prices, between 1896 and 1938, are given in Table 7 and are plotted in Figure 7. Coefficients of the elasticity of lead shipments, with reference to lead prices, are given in Part B of Table 7.

TABLE 7

AVERAGES DEFINING THE PATTERN OF LEAD ORE PRICES IN ELEVEN SPECIFIC CYCLES AND THE PATTERN OF LEAD ORE SHIPMENTS IN THE PRICE FRAMEWORK, 1896–1938*

A. Stage Averages

	(Cycle relatives, price framework)								
	I	II	III	IV	V	VI	VII	VIII	IX
Quantity	78	93	102	114	134	132	110	95	88
Prices	81	87	98	114	127	117	107	91	84

B. Coefficients of Elasticity of Shipments in Price Framework

	Stage Measures Interstage Period							
	I–II	II–III	III–IV	IV–V	V–VI	VI–VII	VII–VIII	VIII–IX
Elasticity	+2.40	+0.78	+0.74	+1.50	+2.18	+0.20	+0.01	+0.06

	Phase Measures Expansion	Contraction	Full Cycle Measure
Elasticity	+1.20	+1.02	+1.11

* Price data relate to the New York market. The quantity series measure shipments in the Joplin District.

We should first note that the framework provided by specific cycles in lead prices does not differ greatly from the general framework of reference cycles, for lead prices conform fairly well to general business cycles.[13] Since the two frameworks are so nearly synchronous we are not able to disentangle the influence of price changes upon lead shipments from the influence of general cyclical forces. Accordingly, although there is a temptation to interpret the relations shown in Figure 7 as those characteristic of a conventional supply function, it would be unsafe to do so. The consistent rise in lead shipments as lead prices rise and the consistent decline in shipments as prices fall may be in part a reflection of a causal relationship running from prices to shipments.

[13] The index of conformity for the full cycle is +82.

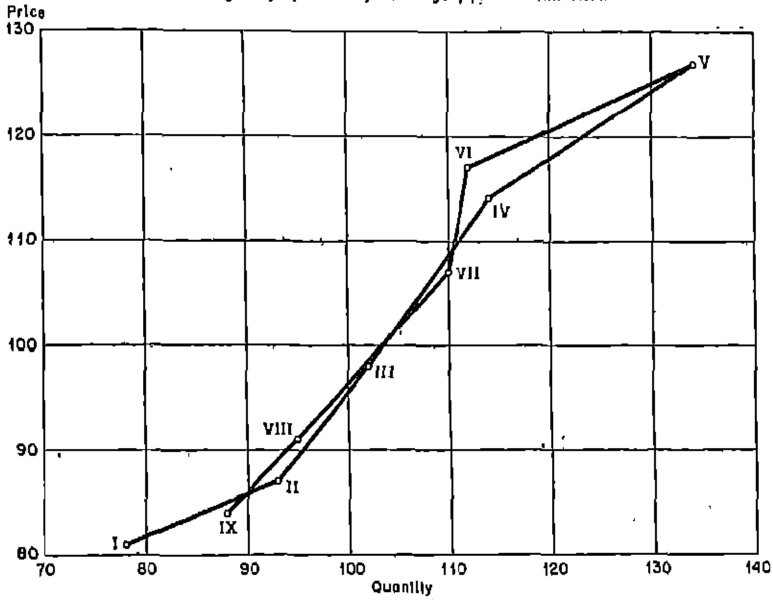
It may also reflect the pressure of common forces, related to cycles in business at large, impinging upon both shipments and prices.[14]

The response of lead ore shipments to price changes is measured by a coefficient of full cycle elasticity of +1.11. The phase coefficient for expansion is somewhat higher than that for contraction, but both

FIGURE 7

Pattern of Prices and Shipments of Lead Ore In Eleven Specific Cycles In Lead Ore Prices, 1896–1938

Averages by specific cycle stages, price framework



exceed unity. The stage coefficients indicate high elasticity of quantities in the first period of expansion and in the first period of contraction, with sharp declines in elasticity in the intermediate stages of expansion and contraction.

V

Elasticity and flexibility during homogeneous secular periods. We pass to relationships between secular movements of economic series. Here we do not seek recurring patterns, as we do in studying cyclical phe-

[14] Although we cannot separate the two sets of influences in this case, we should note that lead shipments conform more closely to specific cycles in lead prices than they do to cycles in general business. The two conformity indexes are +100 and +64, respectively, the former being highly significant, the latter significant. The variance test reenforces this evidence. There is an indication here that the relations shown in Figure 7 represent responses of quantities to independent price changes, although these movements are not isolated from cyclical movements in business at large.

nomena, but regularity that is manifest in a persistent inter-cycle relationship. In Table 8 are given cycle averages for agricultural production in the United States for the seven business cycles occurring between 1911 and 1938, and ratios of the prices received by farmers to prices paid by farmers for goods used in farm production and family maintenance, averaged for the same seven cycles. These averages are

TABLE 8

VOLUME OF AGRICULTURAL PRODUCTION AND RATIO OF PRICES RECEIVED BY FARMERS TO PRICES PAID BY FARMERS*
AVERAGES BY REFERENCE CYCLES, 1911–38

Reference Cycle Terminal Dates (annual)	Agricultural Production†	Ratio of Prices Received by Farmers to Prices Paid by Farmers‡
	(Cycle average)	
1911–14	83.5	98.5
1914–19	86.7	105.4
1919–21	89.5	98.0
1921–24	91.8	82.2
1924–27	98.3	88.3
1927–32	99.6	78.5
1932–38	96.6	76.0

* Data from the United States Bureau of Agricultural Economics.

† Relatives on the base 1935–39.

‡ Relatives on the base August 1909–July 1914.

plotted, on double logarithmic scale, in Figure 8. The averag relaetionship between the price ratios and the production series is defined by the straight line fitted to the observations plotted in Figure 8. The equation to the line is $\log Y = 4.8951 - 1.4995 \log X$, where $\log Y$ and $\log X$ are, respectively, the logarithms of the price ratios and the production indexes.

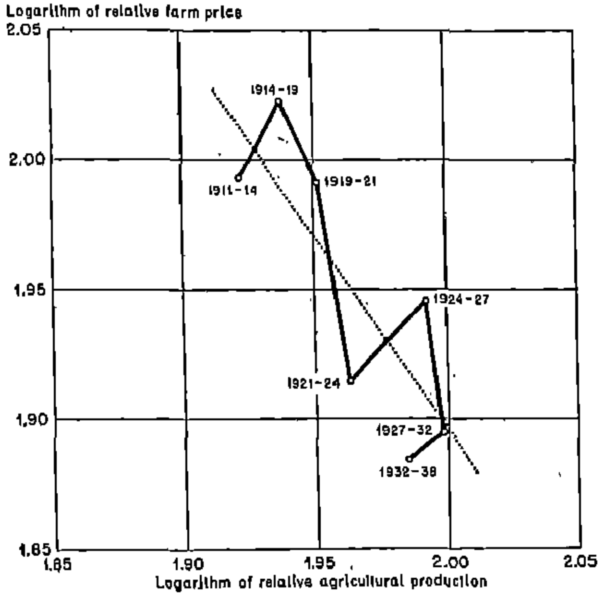
The measure of immediate interest to us is the coefficient of $\log X$ in the above equation. This defines the inter-cycle flexibility of agricultural prices ("real" prices, in the sense that the ratio of prices received to prices paid defines average unit purchasing power of farm products).[15] We interpret the coefficient, which has a value of -1.50, in this manner: Inter-cycle relations between farm output and farm prices were such, over the course of the seven cycles occurring between 1911 and 1938, that for every increase of 1 per cent in the volume of agricultural production there was a decline of 1.50 per cent in the real unit price of farm products. Real farm prices, that is, were flexible, inversely, in response to changes in farm output.

[15] $f = \frac{dy}{dx} \cdot \frac{x}{y} = \frac{d \log y}{d \log x}$.

This relationship prevailed, with notable persistence, over the period of the present records. There were temporary departures from it but it is clear from the graphic record that between 1911 and 1938 rising output was accompanied, secularly, by still more sharply declining real

FIGURE 8
Volume of Agricultural Production and 'Real' Prices Received by Farmers over a Period of Seven Business Cycles 1911—1938

Averages by reference cycles



prices for farm products. The story for recent years is, of course, different, and the record prior to 1911 was marked by other relationships. But for about a quarter century a persistent negative relationship prevailed between farm output and unit purchasing power, and set its mark upon the economic and social conditions of our time.

A final example involves the comparison of inter-cycle movements (i.e., of secular changes) in the volume of manufacturing employment in the United States and output per wage earner in manufacturing plants (Table 9). The data are presented graphically in Figure 9, which is drawn on double logarithmic scale. In this case we are not dealing with unit prices and physical quantities, but the relations in-

volved lend themselves to definition by coefficients of elasticity and flexibility of the type employed in preceding examples.

The secular relations between volume of employment in manufacturing industries and output per wage earner underwent a marked change at about the middle of the period here covered. The averages of employment and productivity are positively related for the first five of the business cycles occurring since the turn of the century. The equation of relationship is $\log Y = -1.3441 + 1.6853 \log X$, where $\log Y$

TABLE 9

MANUFACTURING EMPLOYMENT AND OUTPUT PER WAGE EARNER IN THE UNITED STATES*

AVERAGES BY REFERENCE CYCLES, 1900–38

Reference Cycle Terminal Dates (annual)	Manufacturing Employment†	Output per Wage Earner†
	(Cycle average)	
1900–04	114.6	107.0
1904–08	130.4	113.9
1908–11	130.5	112.5
1911–14	140.5	124.0
1914–19	177.0	135.2
1919–21	177.0	127.5
1921–24	166.7	151.8
1924–27	175.5	171.0
1927–32	161.4	188.0
1932–38	157.2	184.7

* Data from Solomon Fabricant, *Employment in Manufacturing, 1899–1939* (National Bureau of Economic Research, 1942).

† Relatives on the base 1899.

and $\log X$ are, respectively, logarithms of the indexes of employment and per capita output. For the six cycles between 1914–19 and 1932–38 (the 1914–19 cycle, which marks a stage of apparent transition, is included in both sets of calculations) the relationship is negative. The equation is $\log Y = 2.7700 - 0.2466 \log X$.

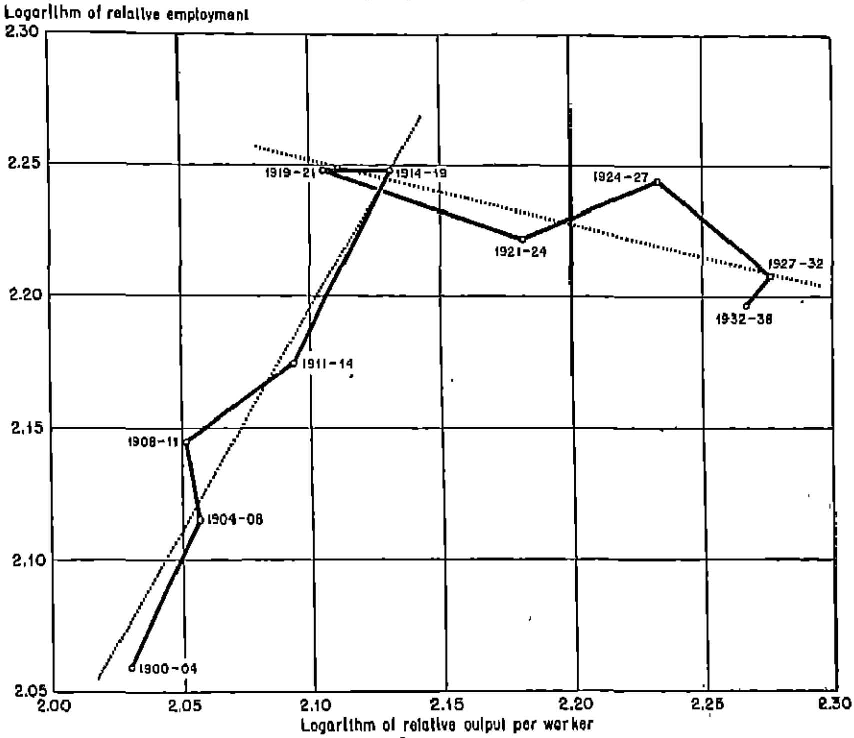
The inter-cycle elasticity of employment, with respect to changes in per capita productivity in manufacturing industries, was +1.69 for the cycles from 1900–04 to 1914–19, −0.25 for the cycles from 1914–19 to 1932–38. That is, during the first of these periods there was a secular increase of 1.69 per cent in volume of manufacturing employment for every advance of 1 per cent in per capita productivity; after 1914–19 there was a secular decline of 0.25 per cent in employment for every advance of 1 per cent in per capita productivity. This is a notable reversal indeed. Advancing productivity was accompanied by in-

creasing employment opportunities in manufacturing during the first two decades of this century. During the second two decades productivity continued to advance, but employment declined. It is true, of course, that the persistent depression of the '30's leaves a marked

FIGURE 9

Manufacturing Employment and Output per Wage Earner over a Period of Ten Business Cycles, 1900–1938

Averages by reference cycles



impress on the cycle averages, but the reversal of earlier relationships seems to antedate the recession that began in 1929.[16]

[16] Such a bare descriptive statement as is provided by a coefficient of employment elasticity with respect to one related variable does not, of course, provide an explanation of the reversal we have noted. Changes in the direction and volume of investment, the increasing importance of service industries, alterations in international economic relations all influenced, directly or indirectly, the volume of employment in manufacturing in the United States. Beyond such specific circumstances, some part of the explanation of the abrupt shift in earlier tendencies, and of the failure of manufacturing employment to advance as industrial efficiency increased, is to be sought in the working of the wage-cost-price mechanisms by which we adapt the use of resources to changing conditions of production and to shifts in demand.

VI

Summary. In the study of demand and price relations emphasis has been placed, traditionally, on the question: What is the elasticity of demand (for a given commodity) with reference to changes in unit prices when no change occurs in any factor other than price that might affect the character of demand? In the present paper we have been concerned with such questions as these: Is quantity (of a given commodity) elastic or inelastic with reference to movements of unit prices within business cycles? Is quantity elastic or inelastic with reference to the trend of prices within relatively homogeneous secular periods? Similar questions might relate to the elasticity of quantities with reference to seasonal movements of prices, movements of prices within long cycles, hour-to-hour movements of prices, or year-to-year movements of prices. In dealing with these questions we deliberately introduce the factor of time, making no attempt to eliminate or hold constant the various forces other than those of volume and price that play upon the market. We seek to measure the differential responsiveness of the quantity and price factors to the various impinging forces that operate within a given temporal framework. In each case the meaning of the coefficient of quantity elasticity (or of the correlative measure of price flexibility) will be determined by the framework within which the price-quantity interactions are studied. The several coefficients for a given commodity may vary widely in value from framework to framework. (Thus the consumption of a given commodity may be inelastic within a seasonal framework, elastic within a framework of business cycles, inelastic with respect to secular changes in unit prices.)

The examples given in the present paper have related, in the main, to quantities and prices, but measurements of the type suggested may be derived for any pair of series whose interaction in a temporal framework is of economic significance. The unit selling prices of manufactured goods and the labor cost of such goods, per unit, may constitute such a pair. The elasticity of physical output with reference to changes in labor costs per unit, and the flexibility of labor costs with reference to changes in output, may be of interest. The relative responsiveness (to cyclical forces, e.g.) of the prices of a given commodity at different distributive stages may be defined by coefficients of flexibility (price flexibility at one stage being measured with reference to price changes at another stage).

Coefficients of the type cited are of limited significance unless they define persistent regularities in economic processes. There may be

persistence of correlated rhythms of prices and quantities in their cyclical or seasonal movements; there may be persistence of trend relations during an extended period. When the existence of continuing uniformities of behavior can be established, appropriately designated coefficients of elasticity and flexibility can serve a highly useful purpose in the study of economic processes.

PRESENTING SEASONAL VARIATION TO THE BUSINESS EXECUTIVE

ALAN S. DONNAHOE
Chamber of Commerce, Richmond, Virginia

In presenting the results of their analysis, business statisticians have failed to reach the full understanding of the average business executive. Even with such a relatively simple concept as seasonal variation, the traditional forms of presentation may be greatly improved and clarified.

STATISTICAL analysis is a mysterious subject to the average business executive. To the extent that results of such analysis are mysterious, they are ineffective. To the same extent, the business statistician has failed in his job.

As a responsible staff member, the statistician should enjoy the full trust of his associates. But trust must be distinguished from credulity, which is defined as "belief on slight or uncertain evidence." If the business executive does not understand the statistical results presented, they definitely represent *uncertain evidence* to him, no matter what his faith in the statistician.

An excellent example may be found in one of the simplest of all statistical operations: adjustment for seasonal variation. This is a subject with which the business executive is intimately acquainted, and yet he is often mystified by the manner in which it is presented by statisticians.

TRADITIONAL FORMS OF PRESENTATION

A brief example will show why the statistician has failed in his presentation of this relatively simple subject. Sales of a hypothetical department store are shown in Table 1. Traditional methods of presenting these data, including the adjusted dollar total and index numbers on a fixed base, are shown in Table 2. Neither method will be clear to the average business executive.

While he may have some sketchy grasp of the concept, the adjusted dollar figures will make little sense to the department store manager. He knows very well that his December volume is greatly in excess of his September volume. He must plan his stock and personnel in precise accord with this difference. To show these at the same general level

is to violate his most basic standards of belief. The index data are even more abstract and difficult to reconcile with his everyday knowledge of sales volume in dollars and cents.

If any statistician doubts that the average executive is confused by this type of presentation, he may carry out a simple experiment with any of the group at hand. Upon seeing such a report, the odds are very great that the excutive will continue his old habit of comparing a given month with the same month of the preceding year. This is his own tried and true method of eliminating seasonal variation and he will not discard it lightly. He naturally arrives at the same result as with unadjusted data. The whole process of adjustment has been worse than useless by adding confusion, but no accuracy, to the usual simple procedure.

When he compares a given month with the same month of a preceding year, the executive realizes that his result is a mixture of unusual conditions that may have prevailed in both periods. While his primary concern is with the current period, this simple method gives him a compound result of the two. It is the statistician who should separate these influences for individual analysis, but this he has failed to do in understandable terms.

TABLE 1. BASIC DATA

ECONOMY DEPARTMENT STORE SALES

In Thousands of Dollars

Month	Total Sales		Seasonal Index
	1945	1946	
Jan.	751	902	90
Feb.	853	1,055	95
Mar.	950	1,105	100
Apr.	1,107	1,154	105
May	1,203	1,213	110
June	1,200	1,177	105
July	1,054	1,010	90
Aug.	960	941	85
Sept.	853	876	80
Oct.	957	1,007	100
Nov.	981	1,203	110
Dec.	1,120	1,431	130
Year	12,000	13,254	100

TABLE 2. TRADITIONAL FORMS OF PRESENTATION

ECONOMY DEPARTMENT STORE SALES

Adj. for Seasonal Variation

Month	Adjusted Sales in Thousands of Dollars		Index: 1935-39 = 100	
	1945	1946	1945	1946
Jan.	834	1,102	139	184
Feb.	808	1,111	160	185
Mar.	950	1,105	158	184
Apr.	1,054	1,099	176	183
May	1,094	1,103	182	184
June	1,151	1,121	192	187
July	1,171	1,122	195	187
Aug.	1,125	1,107	187	184
Sept.	1,066	1,095	178	182
Oct.	957	1,097	159	183
Nov.	892	1,094	149	182
Dec.	866	1,101	144	183
Year	1,005	1,105	167	184

SUGGESTED FORM OF PRESENTATION

One solution to the problem will be found in Table 3, using the same hypothetical data as before. Let us examine this presentation as it would be explained to the executive for whom it is designed.

Column 1. Actual dollar sales for 1945. This offers no difficulty.

Column 2. Dollar sales for 1945 as they would have been if distributed in the usual seasonal pattern. This is easily explained.

Column 3. Actual dollar sales for 1946, which offer no difficulty.

Column 4. Percentage change in actual dollar sales, the usual form of analysis in business.

Column 5. Current year compared with preceding year, in percentage form, with unusual influences ironed out of the latter, and with seasonal influence excluded as well.

The final column is simply the seasonally adjusted index of sales, with the preceding as the base year, in percentage rather than index form. There is no difference except in manner of presentation. The traditional form of presentation is mysterious. The form shown is simple, the mechanism is laid bare, and the meaning is clear.

Our executive can readily understand, indeed he has a very keen appreciation of the fact, that some months in the preceding year were above or below that expected from the total volume of sales for the year. The suggested form of presentation shows the exact extent of that deviation in simple fashion.

TABLE 3. SUGGESTED FORM OF PRESENTATION

ECONOMY DEPARTMENT STORE SALES
In Thousands of Dollars

Month	1945		1946 Actual	Per Cent Change	
	Actual	Normal		Actual	Normal
Jan.	751	900	992	32.1	10.2
Feb.	853	950	1,055	23.7	11.1
Mar.	950	1,000	1,105	10.3	10.5
Apr.	1,107	1,050	1,154	4.2	9.9
May	1,203	1,100	1,213	.8	10.3
June	1,200	1,050	1,177	− 2.0	12.1
July	1,054	900	1,010	− 4.2	12.2
Aug.	956	850	941	− 1.6	10.7
Sept.	853	800	876	2.7	9.5
Oct.	957	1,000	1,097	14.6	9.7
Nov.	981	1,100	1,203	22.6	9.4
Dec.	1,126	1,300	1,431	27.1	10.1
Year	12,000	12,000	13,254	10.5	10.5

The executive is dealing with familiar units in a familiar form. There are no abstract index numbers, no terms such as *adjusted for seasonal variation* with a connotation of mystery.

Finally, he can continue his usual practice of comparing similar periods in different years. This is a process that he clearly understands and readily applies. But there is the vast difference that he can now see in quantitative terms exactly what is involved in this comparison: unusual conditions in the preceding as well as in the current period.

SUMMARY

To present results in a way that will be readily understood and used, the business statistician must reorient his thinking in terms of the user. He must avoid abstract numbers and technical jargon. He must take full advantage of familiar forms of analysis in business, such as the comparison of similar periods in different years.

Statisticians may well afford to give more thought to methods of presentation. What has been attempted here, with a relatively simple concept, should be extended to the more elaborate techniques. More lucid reporting is a real prerequisite to widespread adoption of statistical research in business, a matter of interest to the entire statistical profession.

SOME APPLICATIONS OF MULTIVARIATE ANALYSIS TO ECONOMIC DATA[1]

GERHARD TINTNER
Iowa State College, Ames, Iowa

This essay proposes to introduce the economic statistician to some of the newer methods of multivariate analysis.[2] The emphasis will be on methods of estimation and not on tests of hypotheses. Tests of significance will be indicated where they have been established.

Estimation by means of multivariate analysis presents certain generalizations of the methods of multiple regression, as for instance presented in Ezekiel's[3] book. These analogies will be emphasized later in the course of the discussion of the various procedures. We will deal only with the following methods:

(1) Discriminant analysis: Here we propose to determine linear functions or "indexes" computed from various measurable characteristics of certain data. The data have been classified into two groups. Discriminant analysis tries to establish linear functions of the characteristics which are such that they distinguish most successfully between these groups. This method was invented by R. A. Fisher. A test of significance utilizes earlier work of Harold Hotelling.

(2) Principal components: We try to answer the following question: Is it possible to analyze a set of variables into a more fundamental set of components ("factors") possibly fewer in number? Which portion of the total variance can be accounted for by each component? The best method in this field is due to Harold Hotelling.

(3) Canonical correlation: Assume we have two sets of variables. How can we determine linear combinations ("indexes") of the variables in each set in such a fashion that the correlation between the indexes becomes a maximum? This method is due to Harold Hotelling.

(4) Weighted regression: Assume that we have a set of variables *all* of which are subject to disturbances ("errors"). How can we find a weighted linear regression function which will give us the "best" estimates of the weighted regression coefficients? This method, evidently closely related to classical multiple regression analysis, is in its present form due to Tjalling Koopmans. It can also be used to answer a ques-

[1] The author is greatly obliged to his colleagues, Prof. W. G. Cochran, J. Nordin, O. Brownlee and A. M. Mood for help and criticism with this paper. He is also very much indebted to Prof. H. Hotelling (Columbia) and the following members of the Cowles Commission (Chicago): J. Marschak, T. Koopmans, L. R. Klein and L. Hurwicz. Journal paper No. J-1373 of the Iowa Agricultural Experiment Station, Ames, Iowa, Project No. 730.

[2] A summary of some of the methods is given in: S. S. Wilks: *Mathematical Statistics*, Princeton, 1943, pp. 252 ff.

[3] M. Ezekiel: *Methods of Correlation Analysis*, 2nd ed., New York, 1941.

tion previously raised by Ragnar Frisch: How many linear relationships exist probably between the variables (multicollinearity)?

In what follows we propose to discuss these methods briefly and with a uniform notation. We will try to avoid lengthy mathematical deductions and presentation of numerical methods. These can easily be obtained in the literature which will be quoted below. No effort has been made to give a complete survey of the literature.

Some examples previously given by other authors will be summarized and new examples will also be presented. These examples are supposed to indicate the wide range of problems to which the methods can be applied. It should be remembered that these examples are only tentative applications of the various methods and should be regarded merely as illustrations. It is to be hoped that they will stimulate more extensive applications in the economic field.

The data which we use in our examples are time series. But we have neglected almost entirely this particularity of the data and the difficulties connected with it.[4] This introduces possibly some biases into the tests of significance because of the serial correlations[5] probably existing in the data. The problem of degrees of freedom in economic time series has been treated by H. T. Davis.[6] No use has been made of these and similar methods. Presence of serial correlation makes the estimates inefficient. But the loss of efficiency is not very considerable if the serial correlation is not too large. It should be remembered, however, that the tests of significance, where they are given, may be influenced by existing serial correlation in the variables.

Another obvious shortcoming of the methods presented below is the fact that they all assume essentially linear relationships existing in the population corresponding to the sample. It is to be hoped that this difficulty can be overcome later and that analogous methods will be developed to deal with nonlinear cases. We may, for instance, use squares, cross products and higher powers etc. of the variables.

1. Notation. Throughout this paper we will carry on our argument in terms of the sample, but always with the end in view to establish estimates for the relationships existing in the population corresponding to the sample.

Let $X_{it}(i=1, 2, \cdots, p, t=1, 2, \cdots, N)$ be a set of random variables. The observations in this sample correspond to a normally distributed

[4] See G. Tintner: "The Analysis of Economic Time Series," *Journal of the American Statistical Association*, Vol. 35, 1940, pp. 93 ff.

[5] R. L. Anderson: "Distribution of the Serial Correlation Coefficient," *Annals of Mathematical Statistics*, Vol. 13, 1942, pp. 1 ff.

[6] H. T. Davis: *Analysis of Economic Time Series*, Bloomington, Ind., 1941, pp. 175 ff.

multivariate population. We assume that each of the variables $X_1 \cdots X_p$ has been observed at N points $t=1, 2, \cdots, N$.

Denote by

$$\bar{X}_i = \sum_{t=1}^{N} X_{it}/N \qquad (i = 1, 2, \cdots, p) \tag{1}$$

the sample means of the p variables. Then:

$$x_{it} = \bar{X}_{it} - X_i \qquad (i = 1, 2, \cdots, p, t = 1, 2, \cdots, N) \tag{2}$$

are the deviations from the means. The sums of squares and products are:

$$S_{ij} = \sum_{t=1}^{N} x_{it}x_{jt} \qquad (i, j = 1, 2, \cdots, p); \tag{3}$$

and the sample variances and covariances:

$$a_{ij} = S_{ij}/(N - 1) \qquad (i, j = 1, 2, \cdots, p). \tag{4}$$

The sample correlation coefficient between X_i and X_j is:

$$r_{ij} = a_{ij}/\sqrt{a_{ii}a_{jj}} \qquad (i, j = 1, 2, \cdots, p). \tag{5}$$

Finally the standardized variables, deviations from the sample means expressed in terms of their standard deviation $\sqrt{a_{ii}}$ are:

$$z_{it} = x_{it}/\sqrt{a_{ii}} \qquad (i = 1, 2, \cdots, p, t = 1, 2, \cdots, N). \tag{6}$$

2. Discriminant Analysis. The first method discussed here is the method of discriminant functions introduced into statistics by R. A. Fisher.[7]

The problem to be solved is the following: Assume we have a set of measurements of a number of variables which are classified into two groups. Which linear combination of the various measurements will best discriminate between the two groups?

Assume that we have N normally distributed observations on p variables X_i which we denote by X_{it}. ($i=1, 2, \cdots, p$, $t=1, 2, \cdots N$). Classify these into two groups for $t=1, 2, \cdots, N_1$ and $t=N_1+1, N_1+2, \cdots, N_1+N_2=N$. We define the means in each group:

$$\bar{X}_i^* = \sum_{t=1}^{N_1} X_{it}/N_1; \qquad \bar{X}_i^{**} = \sum_{t=N_1+1}^{N} X_{it}/N_2. \tag{7}$$

[7] R. A. Fisher: "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, Vol. 7, 1936, pp. 179 ff. See also: "The Statistical Utilization of Multiple Measurements," *ibid.*, Vol. 8, 1938, pp. 376 ff. *Statistical Methods for Research Workers*, 8th ed., London 1941, section 49.2, pp. 279 ff.

Let the differences of the means be:

$$d_i = \overline{X}_i^{**} - \overline{X}_i^{*} \qquad (i = 1, 2, \cdots, p). \tag{8}$$

We want to find the linear function of the differences of the means:

$$Z = k_1 d_1 + k_2 d_2 + \cdots + k_p d_p \tag{9}$$

which discriminates most successfully between the two sets of variables; this is to say whose square is a maximum relative to its variance. Maximizing the square of (9) under the condition that its variance is constant we get e.g. the following set of equations for the k_i:

$$\begin{aligned} S_{11}k_1 + S_{12}k_2 + \cdots + S_{1p}k_p &= d_1 \\ S_{12}k_1 + S_{22}k_2 + \cdots + S_{2p}k_p &= d_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ S_{1p}k_1 + S_{2p}k_2 + \cdots + S_{pp}k_p &= d_p \end{aligned} \tag{10}$$

The solutions k_i are *proportional* to the estimates of the coefficients of the linear function which in the population corresponding to the sample discriminates best between the two groups in the sense defined above. The similarity of the system of equations (10) and the normal equations in multiple regression analysis should be noted.

A test of significance has been indicated by R. A. Fisher which makes use of Hotelling's generalized Student distribution.[8] This distribution was already derived in 1931. Define a quantity analogous to the multiple correlation coefficient:

$$R^2 = N_1 N_2 (k_1 d_1 + \cdots + k_p d_p)/N. \tag{11}$$

Then the variance ratio

$$F = \frac{(N - p - 1)R^2}{p(1 - R^2)} \tag{12}$$

has Snedecor's F distribution for $n_1 = p$ and $n_2 = N - p - 1$ degrees of freedom. In this fashion we can test the hypothesis that the empirical discriminant function may have arisen out of pure chance, if in reality there is no difference at all between the variates in the two groups in the population. This test is also related to some work of the Indian school of statistics.[9]

[8] H. Hotelling: "The Generalization of Students Ratio," *Annals of Mathematical Statistics*, Vol. 2, 1931, pp. 360 ff.

[9] See, e.g., P. C. Mahalanobis: "On Generalized Distance in Statistics," *Proceeding of the National Institute of Science of India*, Vol. 12, 1936, pp. 49 ff. R. C. Bose and S. N. Roy: "The Distribution of the Studentized D^2-Statistic," *Sankhya*, Vol. 4, 1938, pp. 19 ff.

Example 1

The method has been applied in a most interesting way by Mr. David Durand to financial data. He utilized it e.g. to discriminate between good and bad loans.[10]

Let X_1 be the down payment, X_2 the price, X_3 the monthly income (all in dollars) and X_4 the length of the contract in months. Then a linear function has been determined which in a sample of 484 good and 485 bad loans, discriminates best between good and bad loans:

$$Z = X_1 - 0.174X_2 + 0.124X_3 - 0.45X_4. \tag{13}$$

This method may also be applied in order to classify various economic phenomena. For instance, a group of prices called sensitive prices was frequently used in an attempt to anticipate more general price movements. The question if a given price should be included into this group could be decided by finding a set of relevant measurements for each of a number of sensitive and non-sensitive commodities and then computing the linear combination of the measurements which discriminates most successfully between sensitive and non-sensitive prices. This discriminant function can then be used in order to classify a given price into one or the other of the two groups.

A similar problem is the classification of prices into prices of consumers' goods and producers' goods which we propose to illustrate by an example.

Example 2

We have tried to apply the methods of discriminant analysis to the following problem: Is it possible to distinguish between the prices of producers' goods and the prices of consumers' goods on the basis of certain measurements connected with their behavior during the business cycle? We are going to use some data collected earlier in a previous book of the author.[11] We will use monthly English wholesale prices, taken from the period 1860–1913. The seasonal and trend have been eliminated from these series by a system of moving averages.

We denote by X_1 the median length of the cycle in months. This is the median of all cycles in the period, measured from minimum to minimum.

X_2 is the median percentage of the duration of cyclically rising prices relative to the total duration of the cycle.

[10] D. Durand: *Risk Elements in Consumer Installment Financing*, Financial Research Program, Studies in Consumer Installment Financing No. 8 National Bureau of Economic Research, New York, 1941, pp. 125 ff.

[11] G. Tintner: *Prices in the Trade Cycle*, Vienna, 1935, Table 2, pp. 110 ff.

X_3 is the median cyclical amplitude expressed as percentage of the trend.

X_4 is the mean monthly rate of change in the cycle (percentage of trend per month).

The fact that the various measurements are given in different units is irrelevant, since the discriminant function is invariant for linear transformations. Our result would not be affected if e.g. X_1 was given in years instead of months.

We will try to construct a kind of "index" which will best discriminate between consumers' goods and producers' goods on the basis of the measures of cyclical behavior indicated above. If we can do this, we would have a method which in a sense would measure most efficiently the "cyclical distance" between prices of various commodities.[13]

The linear discriminant function will be in our case:

$$Z = k_1X_1 + k_2X_2 + k_3X_3 + k_4X_4. \tag{14}$$

TABLE I

CYCLICAL MEASUREMENTS

Price	X_1	X_2	X_3	X_4	Z
		Consumers' Goods			
Rice	72	50	8	0.5	0.186
Tea	66.5	48	15	1.0	0.224
Sugar	54	57	14	1.0	0.200
Flour	67	60	15	0.9	0.228
Coffee	44	57	14	0.3	0.183
Potatoes	41	52	18	1.9	0.207
Butter	34.5	50	4	0.5	0.098
Cheese	34.5	46	8.5	1.0	0.128
Beef	24	54	3	1.2	0.070
Average $\bar{X}_i^*$	48.611	52.667	11.056	0.922	0.170017
		Producers' Goods			
Gasoline	57	57	12.5	0.9	0.104
Lead	100	54	17	0.5	0.203
Pig Iron	100	32	16.5	0.7	0.283
Copper	96.5	65	20.5	0.9	0.315
Zinc	79	51	18	0.9	0.206
Tin	78.5	53	18	1.2	0.266
Rubber	49	50	21	1.6	0.238
Quicksilver	155	44	20.5	1.4	0.404
Copper Sheets	84	64	13	0.8	0.243
Iron Bars	105	35	17	1.8	0.208
Average $\bar{X}_i^{**}$	90.30	50.50	17.40	1.07	0.270083
General Average $\bar{X}_i$	70.553	51.526	14.305	1.000	0.227871
Difference d_i	41.689	−2.167	6.344	0.148	0.100021

[13] H. Hotelling: "Spaces of Statistics and their Metrisation," *Science*, Vol. 47, 1928, pp. 149 ff.

This discriminant function should become a maximum while its variance is constant.

We have chosen 19 prices. The various measurements for the data are indicated in Table 1.

The matrix of the sums of squares and products computed from the data given above is presented in Table 2.

TABLE 2

SUMS OF SQUARES AND PRODUCTS

	X_1	X_2	X_3	X_4
X_1	18,382.450	−1,350.960	1,833.410	21.393
X_2		1,275.349	−45.023	−18.794
X_3			495.146	16.345
X_4				3.400

Only the elements above the diagonal are given since the matrix is symmetrical.

The system of equations to determine our estimates k_i is taken from Table 2 and the d_i from the last line of the previous table:

$$
(15)\qquad \begin{aligned}
18{,}382.450k_1 - 1{,}350.960k_2 + 1{,}833.410k_3 + 21.393k_4 &= 41.689\\
-1{,}350.960k_1 + 1{,}275.349k_2 - 45.023k_3 - 18.794k_4 &= -2.167\\
1{,}833.410k_1 - 45.023k_2 + 495.146k_3 + 16.345k_4 &= 6.344\\
21.393k_1 - 18.794k_2 + 16.345k_3 + 3.400k_4 &= 0.148.
\end{aligned}
$$

The solutions are indicated in the following linear discriminant functions:

$$
(16)\qquad Z = 0.001605X_1 + 0.000277X_2 + 0.000825X_3 + 0.002115X_4.
$$

The meaning of the function (16) is as follows: If Z is larger than at the general mean (0.227871), a commodity should be classified as a producers' good, in the opposite case as a consumers' good. The average Z for producers' goods is 0.279083 and for consumers' goods 0.170017. Only one consumers' good (flour) and one producers' good (gasoline) are misclassified.

The values of Z for various commodities and also for the averages are indicated in the last column of Table 1. It is interesting to note that in this function Z (16) the largest weight has been given to X_3 (amplitude). This seems to indicate that the cyclical amplitude is possibly more important than other characteristics in distinguishing consumers' and producers' goods.

We compute R^2 from the last line of Table 1 as $R^2 = (90)(0.109921)/19 = 0.520678$. The following is the variance ratio: $F = 3.802$. This is

clearly significant. We require at the 5 per cent level of significance for 4 and 14 degrees of freedom an F of only 3.11. But at the 1 per cent level an F of 5.03 is required. The null hypothesis that our discriminant function could have arisen by pure chance is refuted by the test. (This test could also have been made by Hotelling's methods without first computing the discriminant function.) Hence it is likely that in the population some difference probably exists in the cyclical behavior of the two groups. We would conclude that there is an effective linear combination of the cyclical measures indicated above which distinguishes successfully between consumers' and producers' goods on the basis of the data used. It is interesting to note that if another consumers' good—namely pepper, is included into the analysis we do not achieve significant results.

Our result is possibly of some economic importance. It should be interpreted in the light of the obvious limitations of our methods in dealing with this problem: The characteristics indicated in Table 1 are probably not really normally distributed in spite of the fact that the median in large samples tends under certain conditions to be normally distributed. It is also possible that a non-linear combination of the characteristics would be more adequate in our case.

If our results were more trustworthy and also based upon a larger sample covering a longer period we could draw more reliable conclusions. We may still tentatively say that our analysis seems to support to a certain degree the contentions of the majority of business cycle theorists.

3. Principal Components. The method presented here was first devised by Hotelling[13] to deal with a problem appearing in factor analysis in psychology:[14] How can we analyze a group of variables into a set of independent i.e. orthogonal components, called "factors?"

Girshick[15] has shown in an important article that the same method can also be applied to the solution of other problems: We have a set of variates, each of which consists of the sum of a systematic component and an error. How can we find a linear function of the variates which is least subject to the "errors"? Girshick also showed that the principal components method leads to maximum likelihood estimates if the variates are normally distributed.

[13] H. Hotelling: "Analysis of a Complex of Statistical Variables into Principal Components," *Journal of Educational Psychology*, Vol. 24, 1933, pp. 417 ff., 498 ff. See also S. S. Wilks: *Mathematical Statistics, op. cit.*, pp. 252 ff.

[14] K. J. Holzinger and H. H. Harman: *Factor Analysis*, Chicago, 1941.

[15] M. A. Girshick: "Principal Components," *Journal of the American Statistical Association*, Vol. 31, 1936, pp. 519 ff.

Assume that we want to replace a set of standardized variables $z_1 \cdots z_p$ by a more fundamental set of variables $u_1 \cdots u_p$. Let us define:

$$
(17) \qquad
\begin{aligned}
z_1 &= k_{11}u_1 + k_{12}u_2 + \cdots + k_{1p}u_p \\
&\cdots\cdots\cdots\cdots\cdots \\
z_p &= k_{p1}u_1 + k_{p2}u_2 + \cdots + k_{pp}u_p
\end{aligned}
$$

These $u_1 \cdots u_p$ are the principal components. We want the u_i to reproduce the original correlations between the variables z_i.

The expression:

$$
(18) \qquad S_i = k_{1i}^2 + k_{2i}^2 + \cdots + k_{pi}^2 \qquad (i = 1, 2, \cdots, p)
$$

is evidently the contribution of the i-th principal component u_i to the variances of all standardized variables z_i. Maximizing (18) under the condition that the correlations between the original variables ought to be reproduced leads to the system of linear equations for the coefficients k_{ij}:

$$
(19) \qquad
\begin{aligned}
k_{1i} + r_{12}k_{2i} + \cdots + r_{1p}k_{pi} &= \lambda_i k_{1i} \\
\cdots\cdots\cdots\cdots\cdots \\
r_{1p}k_{1i} + r_{2p}k_{2i} + \cdots + k_{pi} &= \lambda_i k_{pi}
\end{aligned}
$$

This system of linear equations is again similar to the normal equations in multiple regression analysis. It can only have a non-trivial solution if the determinant is equal to zero.

$$
(20) \qquad
\begin{vmatrix}
(1-\lambda_i) & r_{12} \cdots r_{1p} \\
\cdots\cdots & \cdots\cdots \\
r_{1p} & r_{2p} \cdots (1-\lambda_i)
\end{vmatrix} = 0.
$$

It can be shown that the *largest* root of (20) is associated with the first principal component which accounts for most of the variance.

Next assume that we have the variables z_i consisting of two parts: The "true" value or mathematical expectation and a random error. We want to find a linear function: $u = k_1z_1 + k_2z_2 + \cdots + k_pz_p$. This ought to be chosen in such a fashion that the variance of the errors is a minimum and that the variance of u is one. Girshick[16] has shown this leads to the previous method. We must choose as λ the largest root of the determinantal equation (20). If we drop the second subscript we get exactly the same solutions as before. This second interpretation may be more useful than the original one in economic data.

[16] *Loc. cit.*, pp. 522 ff.

Girshick[17] has also shown that the method of principal components results from the maximum likelihood approach if the original variates X_{it} follow a normal multivariate distribution. Hence it follows that it provides estimates of the principal components in the population corresponding to the sample which have certain optimum properties associated with maximum likelihood solutions. The results of this method need not, however, be meaningful in economic terms.

The distribution of the latent roots of the determinantal equation (20) has been established by various authors.[18]

Use has been made recently by M. J. Hagood[19] and E. H. Bernert of the method of principal components in the field of sampling of economic data.

The most important class of problems to which it could be applied are perhaps those connected with statistical questions arising from the transition from micro-economic to macro-economic analysis.[20] These questions have been discussed from the point of view of economic theory,[21] but never verified statistically by the use of valid methods.

The practical importance of a solution of this problem lies in the following:[22] Many questions of economic policy require a knowledge of the broad economic relationships which are discussed in economic theory under the name of general equilibrium. This is true for instance of problems of full employment, taxation, subsidies, etc., which ought to be discussed in the most general terms possible. It is obviously impossible to verify statistically a true general equilibrium because of the great number of variables involved. It would be necessary to include all prices of all commodities, all quantities of all commodities produced and consumed, all interest rates, etc. It is obvious that such a procedure would literally involve thousands and possibly millions of variables.

Hence it seems to be necessary to substitute certain indexes for groups of these variables. We may want for instance to represent all wholesale prices by an index of wholesale prices, all quantities produced by an index of production etc. Which particular indexes will

[17] *Loc. cit.*, pp. 527 ff.

[18] M. A. Girshick: "On the Sampling Theory of the Roots of Determinantal Equations," *Annals of Mathematical Statistics*, Vol. 10, 1939, pp. 203 ff. R. A. Fisher: "The Sampling Distribution of Some Statistics Obtained from Non-Linear Equations," *Annals of Eugenics*, Vol. 9, 1939, pp. 238 ff. P. L. Hsu: "On the Distribution of Roots of Certain Determinantal Equations," *Ibid.*, pp. 250 ff. S. S. Wilks: *Op. cit.*, pp. 261 ff. (unpublished results by A. M. Mood).

[19] M. J. Hagood and E. H. Bernert: "Component Indexes as a Basis for Stratifying a Sample," *Journal of the American Statistical Association*, Vol. 40, 1945, pp. 330 ff.

[20] R. Frisch: "Propagation Problems and Impulse Problems in Dynamic Economics," *Economic Essays in Honor of Gustav Cassel*, London, 1933, pp. 171 ff. M. Kalecki: "A Macrodynamic Theory of Business Cycles," *Econometrica*, Vol. 3, 1935, pp. 327 ff.

[21] See, e.g., O. Lange: *Price Flexibility and Employment*, Bloomington, Indiana, 1944, pp. 103 ff.

[22] G. Tintner: "Multiple Regression for Systems of Equations," *Econometrica*, Vol. 14, 1946, pp. 5 ff., esp. pp. 6–9.

be chosen depends of course upon the nature of the economic problem considered. But it is of interest to establish the *statistical* validity of these indexes in the following sense: How perfect is the representation of all the various prices, for instance, by some general price index? Which percentage of the variance of various quantities produced in the economy is accounted for by a certain production index? We believe that questions of this nature can be answered tentatively by the method of principal components.

These problems are evidently connected with the general problem of index numbers.[23] They have very far-reaching significance for the choice between several possible macro-economic models and their empirical validity, if we strive for econometric applications of these models.

The very efficient computational methods developed by Hotelling[24] have been utilized in the following examples:

Example 3

The first example deals with an attempt to determine the principal components of a set of production indexes. This example is somewhat related to an earlier essay of E. C. Rhodes.[25]

Denote by X_1 an index for the production of manufactured durable goods, X_2 of non-durable manufactured goods, X_3 of minerals and X_4 of agricultural products. All indexes are computed with the base 1935–39 = 100. The period covered is 1919–39. We use annual figures. The indexes X_1, X_2, X_3 are taken from the publications of the Federal Reserve Board and X_4 from the year-book of Agricultural statistics of the Department of Agriculture. The correlation matrix is given in the following table:

TABLE 3
CORRELATION MATRIX

	X_1	X_2	X_3	X_4
X_1	1.000000	0.495041	0.872836	0.481240
X_2		1.000000	0.768270	0.709807
X_3			1.000000	0.716359
X_4				1.000000

These four variables can be analyzed into various components or factors.

[23] See, e.g., G. von Haberler: *Der Sinn der Indexzahlen* Tuebingen, 1927. R. Frisch: "Annual Survey of Economic Theory: The Problem of Index Numbers," *Econometrica*, Vol. 4, 1936, pp. 1 ff.

[24] H. Hotelling: "Simplified Calculation of Principal Components," *Psychometrika*, Vol. 1, 1936, pp. 27 ff.

[25] E. C. Rhodes: "The Construction of an Index of Business Activity," *Journal of the Royal Statistical Society*, Vol. 100, 1937, pp. 18 ff.

We want to find the principal component which is such that it contributes most to the variances of the standardized variables $z_1 \cdots z_4$.

The system of linear equations which yields the coefficients of the first and largest principal component is:

$$
(21)\quad
\begin{aligned}
1.000000k_{11} + 0.495941k_{21} + 0.872836k_{31} + 0.481240k_{41} &= \lambda k_{11}\\
0.495941k_{11} + 1.000000k_{21} + 0.768279k_{31} + 0.709807k_{41} &= \lambda k_{21}\\
0.872863k_{11} + 0.768279k_{21} + 1.000000k_{31} + 0.712358k_{41} &= \lambda k_{31}\\
0.481240k_{11} + 0.709807k_{21} + 0.712358k_{31} + 1.000000k_{41} &= \lambda k_{41}
\end{aligned}
$$

This system of equations can only have non-trivial solutions if its determinant becomes zero. The determinantal equation becomes:

$$
(22)\quad
\begin{vmatrix}
1.000000-\lambda & 0.495941 & 0.872836 & 0.481240\\
0.495941 & 1.000000-\lambda & 0.768279 & 0.709807\\
0.872836 & 0.768279 & 1.000000-\lambda & 0.712358\\
0.481240 & 0.709807 & 0.712358 & 1.000000-\lambda
\end{vmatrix} = 0.
$$

From this we have $\lambda = 3.033424$. The contribution of the first principal component to the variance of the standardized variables are the squares of: $k_{11} = 0.817391$, $k_{21} = 0.888102$, $k_{31} = 0.951934$, $k_{41} = 0.818776$. Hence it follows that the first principal component "explains" about 67 per cent of the variance of z_1, about 79 per cent of the variance of z_2, about 91 per cent of the variance of z_3 and about 67 per cent of the variance of z_4.

The same results can also be used to exemplify the other approach. We assume now that each variable z_i consists of a systematic part, the mathematical expectation, and a random component. The function u which minimizes the error variances (while its own variance is one) is:

$$(23)\quad u = 0.269462z_1 + 0.292772z_2 + 0.313815z_3 + 0.269918z_4.$$

The coefficients in (23) are proportional to the previous ones. It is interesting to note that minerals have the greatest weight.

The total variance of the four standardized variables is evidently 4. Hence, since $\lambda = 3.033424$ it appears that the first principal component "explains" about 76 per cent of the total variance of the standardized variates $z_1 \cdots z_4$.

The economic interpretation of these results would appear to be as follows: There existed in all probability during the period considered in the American economy a phenomenon like "production in general." This general "factor" would account for more than $\frac{3}{4}$ of the total vari-

ance of the individual production indexes. A more detailed analysis should of course be carried out to confirm this result not only with respect to the broad categories of production used here but also apply it to the production of individual commodities.

This result, while not unexpected, is by no means trivial. It would be possible, for instance, to imagine an economy where the industrial sector and the agricultural sector have very little relationship. Then we would have two important factors, say, industrial production and agricultural production. The first named would probably account for most of the variance in X_1, X_2, and X_3 and the second for most of the variance of X_4. This is obviously not the case in our example. The factor "production in general" which we have indicated in (23) accounts for most of the variance of X_4 as well as for most of the variance of all other variables.

Example 4

A second example which will not be presented in such great detail deals with prices. Denote by X_5 an index of wholesale farm prices, by X_6 an index of wholesale food prices, by X_7 an index of all other wholesale prices. These are taken from the Bureau of Labor Statistics indexes for the period 1919–1939. The base year of the indexes is 1926. The indexes are given annually. We want again to find the first principal component which accounts for most of the variances of the standardized variates.

An analysis of the data reveals that the contributions of the first principal component to the variance of each standardized variable z_5, z_6, z_7 is the square of the corresponding coefficient: $k_{51}=0.986867$, $k_{61}=0.990160$ and $k_{71}=0.957621$. It appears that the first principal component accounts for about 97 per cent of the variance of z_5, for about 98 per cent of the variance of z_6 and about 92 per cent of the variance of z_7.

The function u which minimizes the variance of the random errors (while its own variance is one) is:

$$u = 0.343693\, z_5 + 0.344845\, z_6 + 0.333508\, z_7. \tag{24}$$

The coefficients in (24) are again proportional to the k's indicated above.

It is remarkable to note that here the weights given to the various variables are approximately the same. The greatest root of the determinantal equation (20) is here $\lambda=2.871360$. Since the total variance of the standardized variables is 3, the principal component can be said to account for more than 95 per cent of the total variance.

It is of some interest to correlate our "index" (24) with the All Commodities Wholesale Price Index computed by the Bureau of Labor Statistics. The resulting correlation coefficient is 0.991 and highly significant for 19 degrees of freedom.

Hence we would conclude that on the basis of the evidence presented it seems that a "general" price index (24) could very well explain most of the variability of the price indexes of groups of commodities. It appears that the residual variability is really almost negligible. This result should, however, be checked by an analysis of the prices of individual goods rather than of broad price categories like those used in our own procedure.

Again this result is what we should expect. But it is by no means as obvious as it seems. We could again for instance imagine an economy in which the industrial and the agricultural sectors have very little connection. Then we would have to distinguish two factors instead of one, say industrial prices and agricultural prices. The first factor would account for most of the variance of X_7 and the second for most of the variances of X_5 and X_6. This is obviously not the case in the American economy. The index indicated in (24) which represents general price movements accounts never for less than 90 per cent of the variance of any among our variables.

4. Canonical Correlations. In economic statistics we desire sometimes to find the relationship between *sets* of variables. The method of canonical correlations, introduced into statistics by Hotelling,[24] provides means for accomplishing this. We replace each of the two sets of variates by a linear combination of the variates contained in each set, the canonical variate. Then we endeavor to maximize the correlation between these two canonical variates, the canonical correlation.

Assume we have p variables X_i and N observations on each variable. The variables are divided into two groups: $i=1, 2, \cdots, p'$ and $i=p'+1, p'+2, \cdots, p$.
We want to find two linear functions:

$$U = k_1X_1 + k_2X_2 + \cdots + k_{p'}X_{p'} \tag{25}$$

and

$$V = k_{p'+1}X_{p'+1} + k_{p'+2}X_{p'+2} + \cdots + k_pX_p. \tag{26}$$

which have maximum correlation with each other. The variances of U and V are supposed to be equal to one. The canonical correlation coefficient between U and V becomes:

[24] H. Hotelling: "Relations between Two Sets of Variables," *Biometrika*, Vol. 28, 1936, pp. 321 ff. See also: S. S. Wilks: *Mathematical Statistics, op. cit.*, pp. 257 ff.

$$R = \sum_{i=1}^{p'} \sum_{j=p'+1}^{p} a_{ij} k_i k_j \tag{27}$$

and this is to be made a maximum under the conditions that the two variances are equal to one.

Maximizing (27) under conditions that the variances of U and V are one we are led to a system of linear equations. These linear equations show again a certain similarity to the normal equations in classical multiple regression analysis. With the help of various transformations which have been given explicitly by Hotelling and need not be repeated here, we are finally led to a determinantal equation:

$$\begin{vmatrix} f_{11} - \lambda^2 & f_{12} \cdots f_{1p'} \\ \cdots & \cdots \\ f_{p'1} & f_{p'2} \cdots f_{p'p'} - \lambda^2 \end{vmatrix} = 0. \tag{28}$$

The f_{ij} are certain functions of the variances and co-variances a_{ij}. Equation (28) determines λ^2, the square of the maximum canonical correlation coefficient. We take the largest root of the determinantal equation (28). The joint distribution of the roots of this equation has been found by various authors,[27] under the hypothesis that their population value is zero. Standard errors have been derived earlier by Hotelling. Inserting the value of λ into the system of linear equations we find $k_1 \cdots k_p$. These provide estimates of the canonical variates in the population corresponding to our sample. These canonical variables are such that U is most successful in predicting V and V the best predictor of U.

It should perhaps be emphasized that these methods do not necessarily yield results which can be readily interpreted in terms of economic theory. This problem will be discussed in greater detail in the last section of this paper.

Hotelling[28] applied canonical correlation first to some pyschological data taken from T. L. Kelley. But he indicated the possibility of applying this method to certain economic problems e.g. the effect of crops of agricultural products on their prices, etc.[29]

The two most successful attempts to apply these methods to economic data have been made by F. V. Waugh.[30] He studied (1) the relation between consumption and prices of various types of mean and(2)

[27] R. A. Fisher: "The Sampling Distribution of some Statistics Obtained from Non-Linear Equations," *Annals of Eugenics*, Vol. 9, 1939, pp. 238 ff. P. L. Hsu: "On the Distribution of Roots of Certain Determinantal Equations," *ibid.*, pp. 250 ff., S. S. Wilks, *op. cit.*, pp. 265 ff.

[28] *Loc. cit.*, pp. 342 ff.

[29] *Loc. cit.*, p. 322, p. 370 f.

[30] F. V. Waugh: "Regressions Between Sets of Variables," *Econometrica*, Vol. 10, 1942, pp. 290 ff.

the relation between characteristics of wheat and characteristics of flour.

Example 5

We indicate the first analysis as follows: Let X_1 be steer prices and X_2 hog prices, X_3 beef consumption and X_4 pork consumption. Then the two canonical variates are: $U=1.71117\ X_1+1.54037\ X_2$ for the prices and $V=5.25679\ X_3+15.45684\ X_4$ for the consumption. These canonical variates are chosen in such a fashion as to maximize the (canonical) *correlation* between U and V. This correlation turns out to be -0.84666. U is the most successful linear combination of the prices to predict V and V is the best linear combination of the consumption data for predicting U.

Example 6

In the other example given by Waugh, let the wheat characteristics be as follows: X_1 kernel texture, X_2 test weight, X_3 damaged kernels, X_4 foreign materials, X_5 crude protein content. The flour characteristics are as follows: X_6 wheat per bbl. of flour, X_7 ash in flour, X_8 crude protein in flour, X_9 gluten quality index. The canonical variate formed from the wheat characteristics is: $U=0.03902\ X_1\ +0.23817\ X_2 -0.03172\ X_3-1.18545\ X_4+0.77554\ X_5$. The canonical variate formed from the flour characteristics is as follows: $V=-0.11971\ X_6-13.12015\ X_7+1.12464\ X_8+0.05903\ X_9$. The (canonical) correlation between U and V is 0.909388. This is the highest possible correlation between any linear combination of wheat and flour characteristics. U may be used to predict V and V is most successful in predicting U.

Example 7

In the following example we will try to determine the relationship between certain price indexes and some production indexes by the method of canonical correlation. The data are the following:

X_1 is the index of production of manufactured durable goods, X_2 of nondurable goods. X_3 is the production index of minerals and X_4 the index for agricultural products. All these indexes are given annually for the base 1935–1939 = 100. They have been taken from the publications of the Federal Reserve Board except for X_4 which comes from the Department of Agriculture. These production indexes form the first group.

The yearly price indexes, all given for the base 1926 = 100 are taken from the publications of the Bureau of Labor Statistics. All are wholesale prices. X_5 denotes farm prices, X_6 food prices and X_7 other prices. The period covered by all these indexes is 1919–1939. They are annual data given for 21 years.

The matrix of the correlation coefficients is represented in the following table:

TABLE 4
CORRELATION MATRIX

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1	1.000000	0.495941	0.872830	0.481240	−0.470785	−0.427250	−0.203300
X_2		1.000000	0.768270	0.709807	0.425728	0.422576	0.644220
X_3			1.000000	0.712358	−0.038273	−0.043762	0.138080
X_4				1.000000	0.261010	0.257098	0.378452
X_5					1.000000	0.087285	0.004509
X_6						1.000000	0.914304
X_7							1.000000

We want to find two linear functions (canonical variates):

$$U = k_1x_1 + k_2x_2 + k_3x_3 + k_4x_4 \tag{29}$$

$$V = k_5x_5 + k_6x_6 + k_7x_7. \tag{30}$$

The variances of these two functions U and V should be one and their correlation a maximum.

Using the iteration methods developed by Hotelling[31] and the computation schemes of Waugh[32] we get the following results:

$$U = 1.004989z_1 - 0.371020z_2 - 0.587650z_3 - 0.020004z_4 \tag{31}$$

$$V = 1.000000z_5 - 0.011424z_6 - 0.215485z_7. \tag{32}$$

These results are given in terms of the standardized variables z_i.

The (canonical) correlation coefficient between these two linear functions of the quantities produced (U) and the prices (V) is 0.8831. U and V are chosen in such a manner that they have the highest possible correlation with each other.

This can also be expressed in the following way: Our U is the linear combination of the various production indexes which is most successful in predicting the general price "index" V. And at the same time V is the linear combination of price indexes which is best in order to predict the general production "index" U.

Needless to say our results are to be interpreted with a certain amount of caution. After all, we are dealing here only with production and price indexes for rather broad categories and not with the production and price data for individual commodities. We can nevertheless reach some tentative conclusions. Our method does not imply that we get necessarily economically meaningful ("structural") relationships.

The first index U (31) shows that in trying to estimate the mutual

[31] *Op. cit.*, pp. 342 ff.
[32] *Op. cit.*, pp. 301 ff.

interdependence between production and prices the largest weight has to be given to the production of *durable* goods. This agrees with the ideas of many students of the business cycle. The weight given to production of minerals is also quite important but negative. Agricultural products seem to play only a very insignificant part. It is especially the weighted *difference* between the movements of the production of durable goods and the production of minerals which appears to be decisive. This points in the direction of certain business cycle theories, especially those stressing the different behavior of various producers' and consumers' goods in the cycle.

Highest weight in V (32) is given to farm prices and some weight also to the index representing all other prices. But this weight is negative. It is probably due to the fact that many miscellaneous prices are contained in this category that the influence is here considerably smaller. Food prices seem not to play a very important part in the determination of V. It is interesting to note that it is the *difference* between farm prices and prices of other commodities which appears to be decisive. This again seems to point in the direction of certain business cycle theories.

5. Weighted Regression. Ordinary multiple regression tries to estimate the relationship between a "dependent" variable and a set of "independent" variables in such a way, as to make the *prediction* of the dependent variable most successful. The sum of squares of the deviations from a linear combination of the fixed values of the independent variables becomes as small as possible (method of least squares). This assumes that we want to predict the dependent variable most successfully for fixed values of the "predictors," i.e. the independent variables.[33]

This method evidently breaks down if we are not interested in prediction but only in the establishment of "structural" relationships existing in the population, and if we also assume that *all* variables are subject to disturbances. For theoretical purposes and also for purposes of economic policy this is most important, as Haavelmo has shown.[34]

Assume we adopt the following stochastic scheme: Not only the dependent variable but *all* variables in the system contemplated are subject to error (Frisch).[35] We do not want to predict one of the variables for fixed values of the others, but want to estimate the structural relationships themselves, i.e. the regression coefficients of the weighted regression equation.

[33] H. Hotelling: "The Selection of Variates for Use in Prediction with Some Comments on the General Problem of Nuisance Parameters," *Annals of Mathematical Statistics*, Vol. 11, 1940, pp. 271 ff.

[34] T. Haavelmo: "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, Vol. 11, 1943, pp. 1 ff. "The Probability Approach in Econometrics," *ibid.*, Vol. 12, 1944, Supplement.

[35] R. Frisch: *Statistical Confluence Analysis by Means of Complete Regression Systems*, Oslo, 1934.

The estimation of structural relationships[36] is most important in connection with problems arising in economic policy. This will be illustrated later especially with the help of Examples 10 and 11. We will, however, indicate here some of the oustanding features of this idea with the help of another illustration.

Consider the market for a commodity, say, wheat. It is known from economic theory that the price of wheat and the quantity of wheat sold on the market are determined by the demand function and the supply function of wheat. If we correlate quantity and price the result need not necessarily represent either the demand function or the supply function. This is irrelevant, however, as long as we assume that the fundamental conditions (tastes, technology, etc.) remain the same and we want only to make predictions. The classical regression of the price on the quantity will for instance under these conditions give the best prediction for the price. And the classical regression of the quantity on the price will be the most successful predictor for the quantity of wheat sold and bought on the market as long as there is no change in the fundamental underlying conditions.

But the situation is entirely different if, for instance, the government decides to fix the price of wheat. Then it becomes most important to know the elasticity of demand. But this elasticity cannot be established from any of the two classical regression equations, except in very exceptional cases.[37] Hence we need a method which will yield estimates of the important economic structural coefficients, like e.g. elasticities, themselves. Weighted regression is designed for this particular purpose rather than for the prediction of values of one particular variable like classical multiple regression.

The problem of distinguishing the various economically meaningful relationships, e.g., demand functions and supply functions, is also very important. This problem of identification will be discussed in more detail in connection with Example 10 below.

The method of weighted regression was developed by Koopmans[38] on the basis of earlier work of many authors, among whom Rodes[39] and van Uven[40] are most important.

[36] A. Wald: "The Fitting of Straight Lines if Both Variables are Subject to Error," *Annals of Mathematical Statistics*, Vol. 11, 1940, pp. 284 ff.

[37] E. J. Working: "What Do Statistical Demand Curves Show," *Quarterly Journal of Economics*, Vol. 41, 1927, pp. 212 ff.

[38] T. Koopmans: *Linear Regression Analysis in Economic Time Series*, Haarlem, 1937.

[39] E. C. Rhodes: "On Lines and Planes of Closest Fit," *Philosophical Magazine*, ser. 7, Vol. 3, 1927, pp. 357 ff.

[40] M. J. van Uven: "Adjustment of N Points (in n-dimensional Space) to the best linear $(n-1)$-dimensional Space," *Koninklijke Akademie van Wetenschapen, Amsterdam Proceedings of the Section of Science*, Vol. 23, 1930, pp. 143 ff., 307 ff.

In the most general form we can pose the problem in the following way: Assume that we have a meaningful (identified)[41] linear relationship between the p economic variables M_i:

$$k_0 + k_1M_1 + k_2M_2 + \cdots + k_pM_p = w \tag{33}$$

where $k_0, k_1, \cdots, k_p$ are "structural" coefficients and w is a random term. It results from variables not included in our system. But actually we don't observe the "true" variables $M_1 \cdots M_p$ but the empirical variables X_{it} $(t=1, 2, \cdots, N)$. We have N observations on each variable. Let us assume that the systematic part M_i is the mathematical expectation of X_i and the y_{it} are the random disturbances.

$$\begin{aligned} X_{1t} &= M_{1t} + y_{1t} \\ &\cdots\cdots\cdots \\ X_{pt} &= M_{pt} + y_{pt} \end{aligned} \qquad (t = 1, 2, \cdots, N). \tag{34}$$

We assume that the "disturbances" or errors y_{it} are independent of each other and normally distributed. They arise as errors of measurement, from lack of representativeness of the empirical variables X_{it}, from frictional causes, etc. It has been proposed to call the y_{it} disturbances in the variables and the w disturbances in the equation.

There are two possibilities in dealing with this situation represented by (33). We can either neglect the disturbances y_{it} or the random term w. This random term results from variables not included in the analysis and similar causes. The first approach is implicit in Haavelmo's,[42] Wald's[43] and Marschak's[44] work. The second assumption underlies the fundamental scheme of Frisch[45] and the weighted regression analysis developed by Koopmans.[46] We are going to deal only with the second case. Our equation (33) becomes:

$$k_0 + k_1M_{1t} + k_2M_{2t} + \cdots + k_pM_{pt} = 0 \quad (t = 1, 2, \cdots, N). \tag{35}$$

We can only neglect w if all or at least most of the important variables have been included in our system.

[41] L. R. Klein: "Pitfalls in the Statistical Determination of the Investment Schedule," *Econometrica*, Vol. 11, 1943, pp. 246 ff.

[42] *Op. cit.* See also: T. Koopmans: "Statistical Estimation of Simultaneous Economic Relations," *Journal of the American Statistical Association*, Vol. 40, 1945, pp. 448 ff.

[43] H. B. Mann and A. Wald: "On the Statistical Treatment of Linear Stochastic Difference Equations," *Econometrica*, Vol. 11, 1943, pp. 173 ff.

[44] J. Marschak and W. H. Andrews: "Random Simultaneous Equations and the Theory of Production," *Econometrica*, Vol. 12 1944, pp. 143 ff.

[45] *Op. cit.* See also: R. Stone: *The Analysis of Market Demand*, London, National Institute of Economic and Social Research, 1945.

[46] *Op. cit.* See also: G. Tintner: "An Application of the Variate Difference Method to Multiple Regression," *Econometrica*, Vol. 12, 1944, pp. 97 ff.

It should be emphasized that it would be most desirable to combine the two methods, i.e. take the disturbances y_{it} into account as well as the random term w. Unfortunately it seems that there are great principal difficulties present.

Assume that we have estimates of the variances of the disturbances y_{it} in the population corresponding to our sample. These estimates V_i may be based upon the Variate Difference Method,[47] upon deviations from time regression functions, e.g. polynomials and Fourier series, or may be given a priori. We could also use the method of principal components as indicated above. The contribution of the first component may be considered as the systematic part and everything else as the random error. We assume that the estimates V_i are accurate enough to enable us to treat them as constants. Then, assuming normality and independence of the disturbances y_{it} we apply the method of maximum likelihood which leads under our assumptions to the method of least squares.

Denote the deviations of the X_{it} and of the M_{it} from their respective means by x_{it} and m_{it}. The weighted sum of squares to be minimized under the condition (35) is then:

$$Q = \sum_{t=1}^{N} \frac{(x_{1t} - m_{1t})^2}{V_1} + \cdots + \frac{(x_{pt} - m_{pt})^2}{V_p}. \tag{36}$$

It should be noted that the weights are the reciprocals of the error variances V_i

It appears that the constant k_0 is determined by the condition that the optimum solution has to pass through the means of all variables:

$$k_0 + k_1\overline{X}_1 + k_2\overline{X}_2 + \cdots + k_p\overline{X}_p = 0. \tag{37}$$

The best solutions for the remaining regression coefficients $k_1 \cdots, k_p$ are given by:

$$\begin{aligned} a_{11}k_1 + a_{12}k_2 + \cdots + a_{1p}k_p &= \lambda V_1 k_1 \\ \cdots\cdots\cdots&\cdots\cdots \\ a_{1p}k_1 + a_{2p}k_2 + \cdots + a_{pp}k_p &= \lambda V_p k_p. \end{aligned} \tag{38}$$

This is a homogeneous linear system in the unknowns $k_1 \cdots k_p$. It is again very similar to the normal equations occurring in classical multiple regression analysis. It can only have a non-trivial solution if its determinant becomes zero:

[47] G. Tintner: *The Variate Difference Method*, Bloomington, Ind., 1940.

$$(39)\qquad \begin{vmatrix} a_{11} - \lambda V_1 & a_{12} \cdots a_{1p} \\ \cdots & \cdots \\ a_{1p} & a_{2p} \cdots a_{pp} - \lambda V_p \end{vmatrix} = 0.$$

The *smallest* latent root of the determinantal equation (39) can be shown to be the minimum of Q divided by $(N-1)$.

By finding the *smallest* root λ_1 of the determinantal equation and inserting it into the previous system (38) we determine the weighted regression coefficients $k_1 \cdots, k_p$.

In equation (35) we can evidently choose one of the $k_1 \cdots k_p$ by an arbitrary condition. Hence putting e.g. $k_1 = -1$ we will change it into:

$$(40)\qquad k_2'm_{2t} + k_3'm_{3t} + \cdots + k_p'm_{pt} = m_1 \qquad (t = 1, 2, \cdots, N).$$

The regression coefficients $k_2' \cdots, k_p'$ are now given by the system of equations:

$$(41)\qquad \begin{aligned} (a_{22} - \lambda_1 V_2)k_2' + a_{23}k_3' + \cdots + a_{2p}k_p' &= a_{12} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{2p}k_2' + a_{3p}k_3' + \cdots + (a_{pp} - \lambda_1 V_p)k_p' &= a_{1p}. \end{aligned}$$

The $k_2' \cdots k_p'$ are our estimates of the weighted regression coefficients in the population corresponding to our sample.

We can use the determinantal equation (39) for a test for collinearity: Let $\lambda_1, \lambda_2, \ldots$ be the *smallest*, next smallest . . . etc. roots of equation (50). We form the test functions:

$$(42)\qquad \Lambda_1 = (N-1)\lambda_1$$

$$(43)\qquad \Lambda_2 = (N-1)(\lambda_1 + \lambda_2).$$

Then it has been shown by Hsu[48] that the test functions Λ_1 and Λ_2 are for large samples distributed like χ^2 with $N-p$ and $2(N-p+1)$ degrees of freedom respectively. If Λ_2 is smaller than the χ^2 computed for a given level of significance (e.g. 5 per cent or 1 per cent), we may conclude that there are probably at least 2 independent relationships between our p variables in the population corresponding to the sample. Hence we have probably collinearity in the sense of Frisch and it is appropriate to fit not one but at least two linear relationships of the type (35). By computing other test functions we can actually estimate the number of independent linear relationships which probably exist in the population corresponding to our sample.

[48] P. L. Hsu: "On the Problem of Rank and the Limiting Distribution of Fisher's Test Function," *Annals of Eugenics*, Vol. 11, 1941, pp. 39 ff. See also: G. Tintner: "A Note on Rank, Multicollinearity and Multiple Regressions," *Annals of Mathematical Statistics*, Vol. 16, 1945, pp. 304 ff.

We can also apply a test of significance for the individual k_i' given by Koopmans which is only approximate.

Denote by c_{ij} the element of the *inverse* of the matrix used to compute the k_i' in (41). This inverse may be computed by the methods given by R. A. Fisher.[49] Then the standard error of the coefficient k_i' which we denote by s_i is given approximately by:

$$(44) \quad s_i^2 = c_{ii}(V_1 + k_2'^2 V_2 + k_3'^2 V_3 + \cdots + k_p'^2 V_p)/(N - p).$$

The ratio k_i'/s_i is approximately distributed like Students' t with $N-p$ degrees of freedom. The t computed in this fashion may also be used to establish fiducial or confidence limits for the weighted regression coefficients. The covariance of k_i' and k_j' may be computed by similar methods. The distribution of the variances and covariances has recently been established for some special cases.[50]

Example 8

The method was applied by Koopmans[51] to the ship freight market, for the period 1880–1911.

Let X_1 be the freight index (1900 = 100), X_2 transport (billions of ton-miles), X_3 tonnage (millions of tons), X_4 coal price (shillings per ton). All variables are expressed in percentages of the trend and the weighted regression equation is: $m_1 = 0.60m_2 + 0.29m_3 + 0.46m_4$.

Example 9

In order to compute these weighted regression coefficients Koopmans had to assume somewhat arbitrarily a set of weights, i.e., the error variances V_i. The author[52] has endeavored to estimate these weights by the Variate Difference Method in a study of agricultural production in the U. S., 1920–1941.

Let X_1 be the logarithm of the volume of agricultural production, X_2 the logarithm of employment in agriculture, X_3 the logarithm of operating capital and X_4 time. The weighted regression equation appears then as: $m_1 = 2.7736\ m_2 + 0.9020\ m_3 + 0.0087\ m_4$. The first two coefficients of this "Douglas type" production function are elasticities with respect to labor and operating capital, while the third coefficient represents an exponential trend. It appears, for instance, that an increase of agricultural employment by 1 per cent will result in an increase of agricultural production by about 2.8 per cent, etc.

[49] R. A. Fisher: *Statistical Methods for Research Workers*, 8th ed., New York, 1941, pp. 150 ff.

[50] T. W. Anderson and M. A. Girshick: "Some Extensions of the Wishart Distribution," *Annals of Mathematical Statistics*, Vol. 15, 1944, pp. 345 ff.

[51] *Op. cit.*, pp. 115 ff.

[52] G. Tintner: "An Application of the Variate Difference Method to Multiple Regression," *loc. cit.*

Example 10

The author[53] has also applied the method of weighted regression in an attempt to find a demand and a supply function for agricultural products in the United States.

Denote by X_1 prices received by farmers for agricultural products, by X_2 national income, by X_3 agricultural production, by X_4 time (origin between 1931 and 1932) and by X_5 prices paid by farmers. The data were given annually for the 24 years 1920–43.

An analysis of the data by weighted regression necessitates the estimation of the error variances. These have again been established by the Variate Difference Method.

An investigation of the problem with the help of the method explained above shows that there are probably *two* relationships between the 5 variables. Other tests show that there is probably one relationship between the variables X_1, X_2, X_3 and X_4 and one between the variables X_1, X_3, X_4 and X_5. The first is evidently the demand function and the second the supply function. It should be noted that the inclusion of X_2 (national income) in the first set and of X_5 (prices paid by farmers) in the second set serves to make the relationships economically meaningful. In this way we *identify* the first weighted regression equation as the demand function and the second as the supply function.

Denoting deviations from the means by m_i we have for the demand function:

$$m_3 = -0.097m_1 + 0.429m_2 + 0.313m_4, \tag{45}$$

and the supply function:

$$m_3 = 1.721m_1 + 0.804m_4 - 3.611m_5. \tag{46}$$

It appears from statistical tests that the results for the equation (45) are more reliable than for (46), as we should expect. Agricultural supply depends largely on the weather and other factors not included in the analysis.

We can compute elasticities from these equations which are based upon the means of the variables over the period. The price elasticity of demand established from the first equation is -0.123. This is to say, other things equal an increase of 1 per cent in the prices of agricultural products results in a decrease of about 0.1 per cent in the quantity demanded. The income elasticity of demand is 0.307. Fiducial or confidence limits can also be established for these elasticities. They appear for the price elasticity of demand as -0.052 and -0.195 at the 5 per

[53] G. Tintner: "Multiple Regression for Systems of Equations," *loc. cit.*

cent level. A test of significance shows that it is highly probable that the income elasticity of demand is definitely greater than the price elasticity. The importance of these tentative results for economic policy are obvious.

Example 11

We propose to illustrate the method of weighted regression further by an example. We will endeavor to fit a production function for the whole American economy in the period 1921–1941, using yearly data for these 21 years. This is mainly an effort to continue the work of Mr. Paul Douglas and his collaborators.[64]

X_1 denotes the logarithm of labor in the U. S. both industrial and agricultural labor in million workers. X_2 is the logarithm of total stock of fixed capital in the economy measured in billions of 1934 \$. X_3 is the logarithm of total private net output also in billions of 1934 \$. X_4 is time measured from 1931 as origin. X_1 is taken from statistical data published by the Department of Agriculture and the Bureau of Labor Statistics. The other data have been taken with kind permission from an unpublished essay by Mr. L. R. Klein.[65]

The means of the data are given in Table 5 below:

TABLE 5

Variable	Symbol	Mean
log labor	X_1	1.661728
log fixed capital	X_2	2.051152
log production	X_3	1.768762

We have indicated above that various methods are available for estimating the variances of the randon elements V_t. We can utilize the Variate Difference Method for this puropse if the following conditions are fulfilled: Each variable consists of the mathematical expectation or systematic part M_{it} which is a smooth function of time, plus the random or error part y_{it}. Then we can eliminate or at least greatly reduce the systematic component by taking differences. If difference series of a high enough order are computed we will eventually have eliminated the systematic part entirely or at least sufficiently. Hence this difference series and all higher difference series will consist of the random part alone, or at least substantially of the random component y_{it}.

In order to form an idea in which difference series this is the case we compute by appropriate formulae the variances of the successive difference series.

[64] P. H. Douglas: *Theory of Wages*, New York, 1934. See also H. T. Davis: *Theory of Econometrics*, Bloomington, Ind., 1941, pp. 153 ff. and the literature quoted on p. 150.

[65] L. R. Klein: *Economic Fluctuations in the U. S., 1921–41*.

We need estimates of the error variances of X_1, X_2, X_3. The following table gives the variances of various difference series of the variables:

TABLE 6

VARIANCES OF DIFFERENCE SERIES

Order of Difference	X_1 log labor	X_2 log capital	X_3 log product
1	0.00026407	0.00001807	0.00082478
2	0.00012103	0.00000250	0.00026453
3	0.00010135	0.00000112	0.00020488
4	0.00000250	0.00000080	0.00018406
5	0.00007757	0.00000065	0.00016423

Tests[56] indicate that the second difference of X_1, the third difference of X_2 and the second difference of X_3 give under our assumptions reasonably accurate estimates of the "true" error variances of the variables in question. The error variances are for the reader's convenience represented in the following table:

TABLE 7

ERROR VARIANCES

Variable	Symbol	Error Variance
log labor	X_1	0.00012103
log capital	X_2	0.000000112
log product	X_3	0.00026453

These error variances are assumed to be estimated with enough accuracy so that we can treat them as constants. This assumption is probably not fully justified in our case.

The resulting weighted regression equation will only represent the production function if, apart from an exponential trend, this production function was stable or at least reasonably stable over the period considered, while there were fluctuations in the other relationship involved in production, especially the supply of productive services, the demand for the product. This assumption seems to be approximately justified because of the fluctuations of those quantities during the business cycle.

We want to fit the weighted regression functions:

$$k_1m_1 + k_2m_2 + k_3m_3 + k_4m_4 = 0 \tag{47}$$

taking into account the fundamental assumptions of our method. The linear equations for the coefficients are derived from the variance-covariance matrix of the variables:

TABLE 8

VARIANCE COVARIANCE MATRIX

$$
\begin{array}{c|cccc|}
 & X_1 & X_2 & X_3 & X_4 \\
X_1 & 0.00003135 & 0.00000045 & 0.00187085 & 0.00515500 \\
X_2 & & 0.00030005 & 0.00045005 & 0.00007500 \\
X_3 & & & 0.00562850 & 0.25020000 \\
X_4 & & & & 38.50000000
\end{array}
$$

[56] G. Tintner: *The Variate Difference Method*, *op. cit.*, pp. 67 ff.

The determinantal equation (29) becomes:

$$
(48)\quad \begin{vmatrix} 0.00093135-0.00012103\lambda & 0.00000045 & 0.00187085 & 0.00515500 \\ 0.00000045 & 0.00030905-0.00000112\lambda & 0.00045005 & 0.00067500 \\ 0.00187085 & 0.00045005 & 0.00562850-0.00020453\lambda & 0.2502000 \\ 0.00515500 & 0.00067500 & 0.25020000 & 38.5000000 \end{vmatrix} = 0
$$

The two smallest roots of this equation are: $\lambda_1 = 0.5482$ and $\lambda_2 = 22.304$. The first can be used to form the test function: $\Lambda_1 = 20\ (0.5482) = 10.9640$. This is for large samples distributed like χ^2 with 17 degrees of freedom. The χ^2 permitted at the 5 per cent level is 27.587 and at the 1 per cent level: 33.409. Hence Λ_1 is not significant.

Next we compute $\Lambda_2 = 20(0.5482 + 22.304) = 458.044$. This is again distributed like χ^2 with 30 degrees of freedom. For the 5 per cent level of significance we get a permissible χ^2 of 50.714 and for the 1 per cent level 57.804. Our empirical Λ^2 is significant. Hence we conclude that it is unlikely that there is more than one linear relationship between the 4 variables X_1, X_2, X_3 and X_4 in the population corresponding to our sample. We do not seem to have multicollinearity.

Inserting the smallest root $\lambda_1 = 0.5482$ into the determinantal equation (48) we get a matrix for the computation of the regression coefficients of our weighted regression: $m_3 = k_1'm_1 + k_2'm_2 + k_4'm_4$.

The equations to be solved are in our case:

$$
\begin{aligned}
& 0.00086500k_1' + 0.00000045k_2' + 0.00515500k_4' = 0.00187085 \\
(49)\quad & 0.00000045k_1' + 0.00030844k_2' + 0.00067500k_4' = 0.00045005 \\
& 0.00515500k_1' + 0.00067500k_2' + 38.5000000k_4' = 0.25020000
\end{aligned}
$$

The solution is given in the following weighted regression equation:

$$
(50)\qquad m_3 = 2.128800m_1 + 0.338605m_2 + 0.005080m_4.
$$

This production function gives estimates of the elasticities of production with respect to labor and fixed capital and also a time trend as they presumably exist in a hypothetical population. For instance, other things equal, an increase of 1 per cent in the total fixed capital will increase the *total product by more than ⅓ per cent. An increase in the total* labor force by 1 per cent will increase the product by more than 2 per cent. The last term represents an exponential trend. It has to be interpreted in this way: Production increased about 3 per cent each year during the period. This last estimate agrees with earlier estimates of Carl Snyder[57] and others.

[57] C. Snyder: *Capitalism the Creator*, New York, 1940.

The weighted regression equation (50) given above should be compared with the regression equation which has been derived by classical methods:

$$x_3 = 1.976985x_1 + 0.332328x_2 + 0.005710x_4. \tag{51}$$

This, however, is designed to predict most successfully x_3 if x_1, x_2 and x_4 are given.

Using the geometric means whose logarithms appear in Table 5, we can also compute the marginal productivities of capital and labor. From the weighted regression coefficients we get for the marginal productivity per worker \$2,279.63 and for the marginal productivity of the stock of fixed capital der pollar: \$0.292. This is to say: If conditions are on the average the same as in the period considered, we conclude that other things equal the addition of one worker will result in an increase in the national product by about \$2,000. The addition of one more dollar to the stock of fixed capital will bring about, ceteris paribus, an increase in the net national product of almost \$.30. Both these estimates appear somewhat high, but maybe not excessively so in the light of some previous investigations in the field of agricultural production functions.[58] These results are, however, not strictly comparable to our production function which has been derived for the whole economy.

All these results should be interpreted in the light of their statistical variability as described by their approximate standard errors. The matrix inverse to the one used in computing our weighted regression equation (5) is given in the following table:

TABLE 9

INVERSE MATRIX

	X_1	X_2	X_4
X_1	1,157.303	41.734	−0.221
X_2		4,700.355	−7.413
X_4			0.038

Using these data and the previous results we compute the approximate standard errors of the weighted regression coefficients. The standard error of the coefficient of m_1 in the weighted regression equation (2.128806) turns out to be 0.174; the one of the coefficient of m_2 (0.338605) appears as 0.351; and the one of the coefficient of m_4 (0,005680) is 0.0009904. Using the t-test, we see that the corresponding values of t are: 12.235, 0,965 and 5.714. The t required for 17 degrees of

[58] G. Tintner and O. H. Brownlee: "Production Functions Derived from Farm Records," *Journal of Farm Economics*, Vol. 26, 1944, pp. 566 ff.

freedom at the 5 per cent level of significance is 2.111 and the one for the 1 per cent level is 2.898. It turns out that the coefficients of m_1 and m_4 are highly significant, but not the one of m_3. Hence it would appear that we can perhaps with some accuracy determine the elasticity of production with respect to labor, but not the one with respect to the stock of fixed business capital. A possible explanation of this is, that the effects of an increase in fixed capital may not appear in the same year but in subsequent years. The time trend can also be determined with reasonable accuracy. All these results are only of an approximate nature.

Finally we want to give fiducial or confidence limits for our estimate of the elasticity of production with respect to labor. Using a confidence coefficient of 99 per cent, we get for the limits of the elasticity: 2.633 and 1.625. This has to be interpreted in the following way: The chances are 99 in 100 that an increase in the total labor force by 1 per cent will increase the product by not more than about 2.6 per cent and not less than about 1.6 per cent. These are pretty wide limits and emphasize the tentative nature of our conclusions.

The same type of analysis can also be applied to the marginal productivity of labor. Using a confidence coefficient of 99 per cent we get for these limits: 2,819.54 and 1,740.13. Ceteris paribus, under conditions approximately the same as the ones prevailing in the period considered, we can make this statement: The chances are 99 in 100 that an increase of the labor force by one worker will result in an increase of the total national product by not more than about \$2800 and not less than about \$1700. The latter figure is probably nearer to the true value.

We want to stress finally that the results for the production function of the whole United States should not be taken too seriously. Our data are perhaps not quite adequate for the determination of such a function. The economic meaning of a production function representing all enterprises is also somewhat doubtful. It would be more desirable to try to fit production functions of the Douglas type to specific industries. We believe, however, that the methods indicated above should be tried in the statistical analysis of such a problem.

REPRODUCTION RATES ADJUSTED FOR AGE, PARITY, FECUNDITY, AND MARRIAGE

P. K. Whelpton
Scripps Foundation for Research in Population Problems

The reproduction rates computed in the past have been age-adjusted, i.e. based on age specific birth rates. Because order of birth and parity of mother were ignored, these rates have had an upward bias in some years and a downward bias in others. The omission of an allowance for marriage and fecundity has had a similar effect. The reasons for these biases are analyzed; a method for utilizing age-parity specific rates and allowing for spinsterhood and sterility is described; and the different types of rates are shown for selected years.

The gross or net reproduction rates and intrinsic rates of natural increase computed in the past have been adjusted for age. They show what would occur if the age specific birth and death rates of females in the various age cohorts of an actual population during the base period were to apply to a hypothetical cohort of females during its lifetime.[1] These age-adjusted rates have been extremely useful to demographers, and have been considered highly accurate measures of the fertility of the base period. As far as the writer can ascertain, however, no one has analyzed adequately the validity of certain phases of the methodology. Is it theoretically possible for the age specific birth rates of any actual population during any base period to remain in effect throughout the life time of a hypothetical cohort? When applying the birth rate of the women of a given age in an actual population to the women of that age in the hypothetical cohort, is it correct to ignore the previous birth rates of the actual women? What is the effect of disregarding the incidence of sterility and spinsterhood? In short, should reproduction rates be based on birth rates which are specific for parity, marriage, and fecundity, as well as age? And if so, how much would they be changed? These are the questions which will be discussed here.

1. CAN A HYPOTHETICAL COHORT HAVE THE AGE SPECIFIC BIRTH RATES OF AN ACTUAL POPULATION?

Heretofore the computation of a gross or net reproduction rate or an intrinsic rate of natural increase has required the use of age specific birth rates. Conventionally, the average annual numbers of live births to white (or colored) women during the base period are classified by

[1] In a few cases reproduction rates have been computed for males and the total population. See, for example, Myers, Robert, J.: The Validity and Significance of Male Net Reproduction Rates. *Journal of the American Statistical Association*, Vol. 36, No. 214, June 1941, pp. 275–282.

5-year age groups (15–19, 20–24, etc.) of mother as of time of birth, corrected for underregistration, and divided by the numbers of white (or colored) women in the corresponding age groups at the middle of the base period. It is assumed that a hypothetical cohort of white (or colored) women living through the childbearing period will have (1) at ages 15 through 19 five times the average annual birth rate which white (or colored) women aged 15–19 had in the base period, (2) at ages 20 through 24 five times the average annual birth rate which white (or colored) women aged 20–24 had in the base period, etc. The total number of births to the hypothetical cohort is computed, and multiplied by the percentage of infants that are girls to obtain the number of female births.[3]

The births by age of mother used in computing the age specific birth rates just described include births of several orders. In 1942, for example, births of the first to sixth orders were registered as occurring to women aged 15–19, and births of the first to 22nd orders to women aged 40–44. Up to the present, however, there apparently has been no reference to order of birth in the methodology of computing a gross or net reproduction rate or an intrinsic rate of natural increase. One explanation of the omission could be that within each age group the rate for first births, the rate for second births, and the rate for births of each other order are assumed implicitly to apply to the hypothetical cohort, just as the sum of these rates (the age specific rate based on births regardless of order) is assumed to apply to it. Another explanation could be that within an age group compensating changes are assumed to occur in rates by birth order. For example, it could be assumed that at ages 20–24 the first birth rates of the hypothetical cohort would be 10 per cent lower than those of the actual cohorts, but the rates for other birth orders would be sufficiently higher so that the total number of births at these ages would not be affected. But because the methodology emphasizes the assumption that the age specific rates (for all birth orders combined) of an actual population apply unchanged to a theoretical cohort—not that there be compensating changes—it is logical to believe that the first assumption mentioned has been made implicitly, namely that the age specific rates by order of birth of an actual population apply to a hypothetical cohort living through the childbearing period.

Is this implicit assumption theoretically possible for any given

[3] Using specific birth rates by single years of age instead of rates by 5-year age groups multiplied by five does not change significantly the gross or net reproduction rate, nor the intrinsic rate of natural increase, unless the distribution of women and/or of births by age of mother within the 5-year age groups is very abnormal. Similarly, using age specific percentages of infants that are girls has little effect on the final results.

year? The answer may be obtained from an inspection of the rates for first births, second births, and births of other orders per 1,000 native white women by 5-year age groups for each year from 1940 to 1944 in the United States. Adding the age specific rates for all births regardless of order and multiplying by five are essential steps in computing the gross reproduction rate, and give the number of births of all orders per 1,000 women living through the childbearing period in the hypothetical cohort.[3] The same procedure applied to first birth rates gives the first births per 1,000 women living to age 50 in the hypothetical cohort. If the cohort had the 1940 age specific first birth rates, 1,000 women would have 820 first births. (See Table 1.) With the 1944 rates there would be 868 first births per 1,000 women, and with the 1941 rates 916 first births. Such fertility is possible in theory or practice. But for 1,000 women to have 1,084 first births, the result obtained for 1942, is impossible both practically and theoretically, even for a hypothetical cohort.[4] And for 1,000 women to have 997 first births (the result obtained for 1943), is equally impossible in view of what is known about the incidence of sterility. Second and higher order births present no such problem in the years for which data are available, because the age specific rates for these birth orders are substantially smaller in the aggregate than those for first births.

TABLE 1

FIRST BIRTHS PER 1,000 NATIVE WHITE WOMEN BY 5-YEAR AGE PERIODS, UNITED STATES, 1940 TO 1944

Age Period	Central First Birth Rates*				
	1940	1941	1942	1943	1944
15–19†	35.09	37.88	42.41	42.47	36.70
20–24	65.88	75.01	90.49	81.09	72.60
25–29	40.87	45.99	55.38	48.24	39.31
30–34	16.41	18.03	21.17	19.88	17.10
35–39	4.82	5.34	6.37	6.74	6.68
40–44	.81	.88	.95	1.00	1.23
45–49‡	.06	.05	.05	.06	.06
Total	163.94	183.18	216.82	199.48	173.68
Total ×5	819.70	915.90	1,084.10	997.40	868.40

* These rates are adjusted for incomplete registration in accordance with data of the Division of Vital Statistics based on the test period December 1, 1939 to April 1, 1940.

† Includes the few births to women younger than 15.

‡ Includes the few births to women 45 or older.

[3] In computing a gross reproduction rate it is assumed that if the women who die before the end of the childbearing period had lived they would have the same age specific birth rates as those who live to the end of the period.

[4] In this connection it should be remembered that multiple live births are recorded and tabulated as births of two or more orders. Thus if twins are borne at a first pregnancy, order of birth is noted as first on the birth certificate for one baby and as second on that for the other.

The conclusion may be stated in general terms as follows: If the various cohorts composing an actual population have a sufficiently large number of first births in a certain year (or years), it is invalid to assume in computing reproduction rates[5] that a hypothetical cohort will have during its reproductive life time the age specific rates for first births of the various cohorts during that year (or years). Furthermore, the fact that the foregoing assumption yields impossible results for certain years casts doubts on the accuracy of the results for other years. Fortunately the situation can be improved in one of three ways. (a) The second assumption listed above can be made, namely, that the age specific rates for first births will be smaller for the hypothetical cohort than for the actual population, but these declines will be exactly balanced as far as numbers of births are concerned by larger age specific rates for births of higher orders. Obviously this is objectionable for several reasons. (b) The base period can be lengthened in the hope that the peculiarities of individual years will cancel out. This is undesirable because it is important to measure year to year changes as well as the average situation during several years. (c) Birth rates can be computed for the actual population which can be applied correctly to a hypothetical cohort. This suggestion seems most promising, hence it will be analyzed in detail.

The first revision in procedure that is needed in order to improve the accuracy of reproduction rates for any year is the inclusion of an adjustment for order of birth of child and parity of women.[6] From a theoretical standpoint the use of birth rates which are specific for birth order and parity as well as age is a decided improvement, and should have been advocated long ago by demographers. The delay probably is due in part to a tendency to accept as adequate in fertility analysis the procedures developed and used earlier in mortality analysis. Unfortunately three fundamental differences have been overlooked in this carry-over.[7] First, although a cat proverbially has nine lives, each woman (no matter how "catty" she may be) has only one; she must die, but she can die only once. In contrast, the fecundity and fertility of women vary widely. Some cannot have a child; many have two or three; a very few have 20 or more. Because of this difference, age specific birth rates are much less adequate in preparing a reproduction table than are age specific death rates in preparing a life table. To raise the

[5] For the sake of brevity the phrase "reproduction rates" will be used hereafter in referring to the three rates—gross reproduction rates, net reproduction rates, and intrinsic rates of natural increase—as a group.

[6] Parity as used in this paper means the number of children born alive. A zero parity woman has not borne a child alive, a first parity woman has had one live birth, etc.

[7] The other two differences relate to sterility and spinsterhood, and are discussed in sections 3 and 4.

former to the level of the latter requires (among other things) recognition of the relation between parity of woman and birth order of child. First births can occur only to zero parity women, second births can occur only to first parity women, etc. In short, only the women of n parity can be exposed to the risk of bearing a child of $n+1$ order.

From a practical standpoint the use of age specific birth rates by order of birth and parity of women[8] for the United States was impossible a few years ago because there were so few statistics regarding the distribution of women by parity. Fortunately, the parity question was reinstated in the 1940 census, and tabulations by parity were made on a sample basis for this census and that of 1910. Tables are now available which show the number of women classified by color, nativity, age, and number of children ever born for the United States and certain subdivisions.

If the age-parity specific birth rates of an actual population are applied to a hypothetical cohort, it is impossible for a gross reproduction rate to show that 1,000 women have more than 1,000 first births. The explanation is simple. First births at each age of mother in the base year are related to zero parity women of that age, and the resulting rates are applied to the zero parity women in the hypothetical cohort. At the beginning of the childbearing period all of the women in the cohort are of zero parity. As these zero parity women have a first birth at any subsequent age they are transferred to the first parity group. If the age specific zero parity first birth rates are sufficiently high, all the women in a hypothetical cohort are transferred from zero to first parity before reaching the oldest childbearing age, in which case the upper limit of 1,000 first births per 1,000 women in the cohort is reached, but not passed. In no year for which data are available, however, has this limit been approached closely. Even with the age specific zero parity birth rates of 1942, the highest on record, 1,000 women living to age 50 would have only 875 first births. Similar reasoning shows that the number of second births in the hypothetical cohort cannot exceed the number of first parity women, or the number of third births exceed the number of second parity women, etc., when the base rates are age-parity specific. Actually, the theoretical upper limits are not approached closely.

2. ALLOWING FOR THE RELATION BETWEEN THE FERTILITY OF A COHORT AT A GIVEN AGE AND ITS FERTILITY AT YOUNGER AGES

The use of age-parity specific birth rates in computing reproduction rates avoids impossible assumptions for years like 1942 because it pro-

[8] Such rates will be referred to hereafter as age-parity specific rates.

vides for the relation between (a) birth performance at a given age and (b) birth performance at younger ages. As was pointed out above, transferring women from zero parity to first parity when they have a first birth and using birth rates by parity make it impossible to assign more than one first birth to any woman. This is desirable for obvious reasons. Less obvious, but important nevertheless, is the fact that because age-parity specific birth rates for a given year give proper weight to the fertility in prior years of each age cohort in the actual population, they measure more accurately than age specific rates the birth performance of the population in the year in question. For example, during the depression years of the 1930's there was a tendency for marriage and the starting of a family to be delayed, which raised the proportion of women who were childless to a relatively high figure, especially at the younger childbearing ages. In contrast, prosperity and war has encouraged marriage and the starting of a family between 1940 and 1945, and reduced to a relatively low figure the proportion of the younger women who are childless. If a comparison of the fertility of the younger childless women in 1940 and 1945 is based on (a) age specific first birth rates and (b) age-parity specific first birth rates, the relative position of 1945 is more favorable on the latter basis than on the former. In this case, of course, the comparison more favorable to 1945 is correct because it reflects the differences in the relative number of zero parity women in the two base years. For the same reason a comparison between the reproduction rates for 1940 and 1945 is more accurate if computed from age-parity specific birth rates than from age specific rates. It is possible in theory for the bias mentioned to be equalized exactly by biases in the opposite direction for other ages or other birth orders. That exact equalization will happen in practice, however, is unlikely.

Basing reproduction rates on age-parity specific birth rates has the additional advantage of allowing for the fact that the computed birth performance prior to a given age of women in a hypothetical cohort will differ in most cases from the actual birth performance prior to the same age of the women in the actual population whose rates are being applied to the cohort. This is important because, as explained above, fertility at a given age is affected by fertility at younger ages. For example, seventh or higher order births constituted more than half of the births in 1940 to native white women aged 40-44, and almost all of them occurred to mothers who had borne at least six children before 1940.[9] In computing a conventional 1940 reproduction rate it is assumed that the women in the hypothetical cohort will have at ages

[9] The few exceptions are the women who had multiple births or two confinements, or both, in 1940.

40–44 the 1940 age specific birth rates at these ages. But whereas the proportion of the women aged 40–44 that had borne six or more children was 13 per cent for the actual 1940 population it would be only 6 per cent for the hypothetical cohort. In consequence the number of seventh or higher order births to 1,000 women in the hypothetical cohort at ages 40–44 is substantially lower, and correspondingly more accurate, when computed from age-parity specific rates than when computed from age specific rates. It might be thought that the smaller number of seventh or higher order births yielded by parity specific rates at ages 40–44 would be offset by a larger number of first to sixth births, because of the larger proportion of zero to five parity women in the hypothetical cohort than in the 1940 population. Such is not the case. Instead, the total number of births per 1,000 women aged 40–44 is 14.9 for the actual population but only 10.7 for the hypothetical cohort. Obviously the reproduction rates based on age specific birth rates are biased accordingly. As before, such biases could happen to be exactly compensating, but the odds are heavily against it.

3. ALLOWING FOR THE EFFECT OF STERILITY

The basic age specific birth rates used in computing conventional reproduction rates are obtained by relating births to the fecund plus the sterile persons in the total population (or in the race, nativity, sex, or other group in question) rather than to the fecund persons only.[10] Failure to exclude sterile persons may be due in part to the tendency suggested earlier to carry over to fertility analysis the procedures used previously in mortality analysis. Relating births, like deaths to all persons is objectionable in theory, however, because it ignores the fact that whereas each person must die sooner or later, some persons in the so-called reproductive age groups cannot become parents. Speaking in terms of women, no women of any age can avoid exposure to the risk of dying at that age. In contrast, some women of an age within the childbearing period cannot, for physiological reasons, be exposed to the risk of having a child at that age. Deaths of women of a given age should be related to all women of that age, for all of them are at risk. Similarly, n order births at a given age should be related to the women who can have such a birth, namely, the *fecund* $n-1$ parity women of that age, and not these women plus the $n-1$ parity women of the same age who are sterile. In discussing this problem a distinction

[10] Fecund and sterile are used in accordance with the definitions adopted by the Population Association of America, namely: a fecund person has the physiological ability to participate in reproduction at the age or time in question; a sterile person lacks this ability.

This analysis deals with one method of allowing for sterility. If age specific rates for the onset of sterility were available, the analysis and conclusions would be somewhat different.

will be made between full time or complete sterility—the lack of the ability to participate in reproduction at any time—and part time or partial sterility—the lack of the ability at one time but not at another. The reasons for this distinction will become evident during the discussion.

From a practical standpoint there are two explanations for not allowing heretofore for the effect of complete sterility in computing reproduction rates. The first is the lack of information about the number of sterile persons in the population, and will be considered later. The second is that a reproduction rate computed from age specific birth rates is not affected by an allowance for sterility. For example, the net reproduction rate for native white women in 1942 based on age specific birth rates with no allowance for sterility is 116. (See Table 2,

TABLE 2

THE NET REPRODUCTION RATE FOR NATIVE WHITE WOMEN IN 1942 COMPUTED FROM AGE SPECIFIC BIRTH RATES, WITH AND WITHOUT AN ADJUSTMENT FOR STERILITY

Exact Ages	Number of women in hypothetical cohort* A	Births per 1,000 women in actual population† B	Per cent of women assumed fecund‡ C	Number of fecund women in hypothetical cohort ($A\times C$) D	Births per 1,000 fecund women in actual population† E	Births to women in hypothetical cohort ($A\times B$) or ($D\times E$) F
15–19	473,059	52.60	90	425,663	58.44	24,930
20–24	470,869	162.50	90	423,782	180.56	76,516
25–29	466,865	144.04	90	420,178	160.04	67,247
30–34	462,031	91.51	90	415,828	101.68	42,280
35–39	455,983	48.11	90	410,385	53.46	21,937
40–44	448,110	14.42	90	403,299	16.02	6,462
45–49	437,429	1.32	90	393,686	1.47	577
Total						239,949
Net Reproduction Rate						116

* l_x values from a life table for 1942 computed by Scripps Foundation for Research in Population Problems.

† See footnotes of Table 1. The figures in these columns are probabilities rather than central rates, for they are obtained by relating (a) the number of births occurring during the age interval x to $x+1$ to (b) the number of women of exact age x.

‡ Any percentage which does not give a rate of over 1,000 in Column E could be used for illustrative purposes.

Col. F.) If it is assumed that 10 per cent of the women are completely sterile, fecund women will constitute 90 per cent of all women of each age in the hypothetical cohort. (See Cols. A and D.) But for the same reason, fecund women will constitute 90 per cent of all women of each age in the actual population, hence the age specific birth rates for fecund women in the actual population (Col. E) will be 111.1 per cent

of those for all women (Col. B). In consequence the number of births to the women in the hypothetical cohort (Col. A×Col. B, or Col. D×Col. E) is the same in either case, as is the net reproduction rate. Treating each single year of age separately does not affect the comparison.

If a reproduction rate is based on age-parity specific rates, however, an allowance for complete sterility changes the results. This may be illustrated by using the fertility and mortality data for native white women in 1942 and assuming that either none or 10 per cent of the women are completely sterile, as before, but computing the net reproduction rates from age-parity specific birth rates. Because these rates must be used by single years of age a table including all birth orders at all ages would require a large amount of space. In consequence, the results are shown in detail only for first births at ages 15 through 24. (See Table 3.) If no allowance is made for sterility, a hypothetical cohort of 100,000 women will have 57,774 first births by exact age 25, and 82,400 by age 50. But if it is assumed that 10 per cent of the women are completely sterile, 100,000 women in the hypothetical cohort will have 57,073 first births by exact age 25, and 79,200 by age 50. Births of other orders will be reduced in similar degree.

The reason why an allowance for complete sterility does not affect a reproduction rate based on age specific birth rates but does affect one based on age-parity specific rates may be explained briefly. In the former case it was pointed out that the relative allowance is exactly the same at each age in the actual population and the hypothetical cohort, and is exactly offset by higher birth rates. In the latter case the relative allowance at each age is exactly the same for *all* women in the actual population and in the hypothetical cohort but *not* for the *zero parity* women. For example, the assumption that 10 per cent of the women are completely sterile at each age results in a group of such women amounting at age 20 to 12.1 per cent of the zero parity women in the 1942 population and to 12.6 per cent of the zero parity women in the hypothetical cohort. At older ages the percentages are less similar, being 19.4 and 25.3 respectively at age 25, and 42.2 and 62.7 at age 40. Such differences occur only when (a) the zero parity women of a given age in an actual population have had at younger ages (and prior to the base year) age specific birth rates which differ from those of (b) the zero parity women of younger ages in the actual population during the given base year (which are also those of the hypothetical cohort prior to the given age). Since it is extremely improbable that the rates in (a) and (b) will be identical, or that the differences will be exactly

TABLE 3

PART OF THE COMPUTATION OF THE NET REPRODUCTION RATE FOR NATIVE WHITE WOMEN IN 1942 FROM AGE-PARITY SPECIFIC BIRTH RATES

With and Without an Adjustment for Sterility and Spinsterhood

Exact age	Number of women in hypothetical cohort*	Number of Zero Parity Women: Total†	Number of Zero Parity Women: Cannot marry and/or are sterile	Number of Zero Parity Women: Can marry and are fecund	Probability of having a first birth ‡ (1,000 p_x^1)	Number of first births ($D \times E$)	Number of women having a first birth and living to next age§
	A	B	C	D	E	F	G
	Assuming All Women Are Fecund and Can Marry before Age 50						
15	94,987	94,987	None	94,987	6.89	653	646
16	94,890	94,253	None	94,253	19.07	1,797	1,789
17	94,801	92,369	None	92,369	41.83	3,863	3,752
18	94,694	88,416	None	88,416	69.31	6,128	6,115
19	94,578	82,198	None	82,198	86.98	7,150	7,137
20	94,454	74,959	None	74,959	120.44	9,028	9,013
21	94,322	65,846	None	65,846	127.46	8,393	8,379
22	94,181	57,373	None	57,373	138.19	7,928	7,914
23	94,033	49,372	None	49,372	150.07	7,409	7,396
24	93,870	41,897	None	41,897	129.52	5,426	5,416
Total	—	—	—	—	—	57,774	—
	Assuming 10 Per Cent of Women Are Sterile, and All Can Marry before Age 50¶						
15	94,987	94,987	9,499	85,488	7.62	651	645
16	94,890	94,264	9,490	84,764	21.20	1,797	1,789
17	94,801	92,370	9,480	82,890	46.58	3,861	3,850
18	94,694	88,419	9,469	78,950	77.53	6,121	6,108
19	94,578	82,208	9,458	72,750	98.04	7,132	7,120
20	94,454	74,986	9,445	65,541	136.09	8,078	8,063
21	94,322	65,923	9,432	56,491	147.09	8,309	8,295
22	94,181	57,534	9,418	48,116	162.37	7,813	7,799
23	94,033	49,648	9,403	40,245	180.02	7,245	7,232
24	93,870	42,335	9,388	32,947	156.80	5,166	5,156
Total	—	—	—	—	—	57,073	—
	Assuming 10 Per Cent of Women Are Sterile and 10 Per Cent Cannot Marry before Age 50**						
15	94,987	94,987	18,049	76,939	8.47	652	646
16	94,890	94,253	18,031	76,222	23.58	1,797	1,789
17	94,801	92,369	18,012	74,357	51.90	3,859	3,848
18	94,694	88,420	17,992	70,428	86.79	6,112	6,099
19	94,578	82,218	17,970	64,248	110.71	7,113	7,101
20	94,454	75,016	17,946	57,069	156.33	8,022	8,007
21	94,322	66,008	17,921	48,087	170.76	8,211	8,198
22	94,181	57,716	17,894	39,822	192.71	7,674	7,660
23	94,033	49,968	17,866	32,102	219.43	7,044	7,031
24	93,870	42,856	17,837	25,019	193.49	4,841	4,832
Total	—	—	—	—	—	60,225	—

* l_x values from a life table computed by Scripps Foundation for Research in Population Problems.

† At age 15 Column B equals Column A. At each subsequent age the figure for Column B is obtained by multiplying the number of single and ever married zero parity women of the preceding age by their respective survival probabilities, and deducting the women in Column G for the preceding age.

compensating, it is equally improbable that an allowance for complete sterility will not change a reproduction rate based on age-parity specific birth rates.

Among the women who are sterile during part of the childbearing period and fecund during the remainder of the period, the fecund period precedes the sterile period in almost all cases, hence the discussion of partial sterility will be based on this group. For reasons explained in connection with full time sterility, an allowance for part time sterility has no effect on a reproduction rate computed from age specific birth rates. Whether it has an effect if the base rates are age-parity specific depends on the type of allowance. Part time sterility undoubtedly increases with age because of the longer period during which there may be exposure to the factors causing fecund women to become sterile, e.g., gonorrheal infection and tumors of the generative organs. It so happens, however, that if the onset of sterility depends entirely on age, an allowance for part time sterility has no effect on reproduction rates computed from either age specific or age-parity specific birth rates. Excluding the partially sterile women reduces each age and parity group in the same proportion in the actual population as in the hypothetical cohort. In consequence the increases in the age specific or age-parity specific birth rates computed for the actual population exactly offset the decreases in the number of fecund women in the hypothetical cohort.

An allowance for a relation between bearing a child and becoming sterile would have no effect on reproduction rates based on age specific birth rates, but would be almost certain to affect reproduction rates based on age-parity specific birth rates. The reason is that (as was pointed out earlier) the parity distribution at a given age of the women in an actual population is almost certain to differ from that of the women in a hypothetical cohort exposed to the rates for the actual population in question. The important question thus becomes: Is there a relation between bearing a child and becoming sterile, and if so what is it? Everyone knows that some of the women who become sterile during a given year of age do so only because of events associated with childbearing. It is less generally realized, but true nevertheless,

‡ In computing a net reproduction rate it is assumed that the women who die before the end of the childbearing period have while living the same age specific birth rates as those who live to the end of the period.

§ An allowance is made at each age for the higher mortality of women who bear a child at that age than of other women.

¶ In this section of the table Column C equals 10 per cent of Column A.

** In this section of the table Column C equals 10 per cent of Column A. This percentage is obtained by multiplying the assumed percentage fecund (00) by the assumed percentage that can marry (00) and subtracting the product from 100.

that the onset of sterility in some of the other women of that age might have been prevented by pregnancy and childbirth. The present difficulty is to estimate the extent to which childbearing causes or prevents sterility. Until more information is available on this matter it is not worthwhile to allow for partial sterility in computing reproduction rates.

With complete sterility, in contrast, the choice lies between disregarding its effects, and estimating its incidence from the best material available. There is no perfect basis at present, nor is there likely to be one in the foreseeable future, for subdividing all women of the childbearing ages into two groups, one always sterile and the other fecund part or all of the time. It is possible, however, to set an upper limit on the proportion completely sterile, and thus determine the range within which the reproduction rates adjusted for this type of sterility should fall. Census tabulations show that of the native white married women aged 45–49 for whom the number of children ever born was reported, 15.7 per cent were of zero parity in 1940 and 10.0 per cent in 1910. Allowing for the underreporting of children ever born reduces the 1940 figure to 12.7 per cent,[11] and probably would reduce the 1910 figure to less than 8.0 per cent. Obviously not all of these women were childless because they themselves were sterile at all times. Some were childless because they were married to sterile husbands, and others, because they and their husbands were of low fecundity. In this analysis, however, such women will be included with those completely sterile, because they have the same effect on reproduction rates. Finally some of the childlessness was due to partial sterility, some to broken marriages, and some to contraception.

Part of the increase in the percentage childless from 1910 to 1940 probably was due to an increase in complete sterility. Studies of human fertility indicate, however, that the greater part was due to the more widespread and effective use of contraceptives. For this reason, it is unlikely that the proportion of native white married women aged 45–49 in 1940 who were always sterile exceeded the proportion reported childless in 1910, namely 10 per cent.

Time will tell what proportion of the women under 45 in 1940 who marry by 45–49 will be childless when they reach that age. In view of the increase in childlessness between 1910 and 1940, however, it probably will exceed 12.7 per cent, the adjusted 1940 figure. Here again it is probable that most of the increase will be due to the more widespread and effective use of contraceptives and relatively little to the greater incidence of complete sterility. It seems safe to conclude,

[11] Unpublished study by the writer.

therefore, that 10 per cent is a maximum estimate for complete sterility among women under 50 in 1940 who will marry before reaching that age. Little is known as to the incidence of complete sterility among women who do not marry before 50. The best that can be done is to assume that there is no relation between complete sterility and marriage, hence that 10 per cent is the maximum estimate for these women also.

4. ALLOWING FOR THE EFFECT OF SPINSTERHOOD

If all children were borne by ever married women and none by single women the reasons for, and the effect of, allowing for spinsterhood would be the same as for sterility. Under these conditions it would make no difference whether age specific birth rates based on all women in the actual population were applied to all women in the hypothetical cohort, or whether age specific birth rates for women who marry in the actual population were applied to women who marry in the hypothetical cohort, provided that the proportion of single women who marry was the same at each age in the hypothetical cohort as in the actual population.[12] If the basic birth rates are specific for parity as well as age, however, the situation is as different for spinsterhood as it is for full time sterility. If all women marry and 10 per cent are completely sterile, the highest conceivable age-parity specific rates could not give more than 900 first births to a hypothetical cohort of 1,000 women living to age 50. But if 10 per cent of the women remain spinsters (and virgins) and 10 per cent of those who marry are completely sterile, the corresponding upper limit for first births is reduced to 810. Within these upper limits the effect of allowing for spinsterhood depends on the extent to which the proportion of the women of each age in the actual population who are of zero parity differs from the corresponding proportion for the hypothetical cohort. Under mortality and fertility conditions of 1942, and assuming that 10 per cent of the women are completely sterile, a hypothetical cohort of 100,000 women will have 57,073 first births by exact age 25 and 79,200 by exact age 50. But if it is assumed also that 10 per cent of the women cannot marry, the hypothetical cohort will have only 56,225 first births by exact age 25 and 75,100 by exact age 50. (See Table 3.) The principle involved was discussed in connection with the allowance for full time sterility and needs no further attention here.

If spinsterhood and virginity were synonymous the ideal way to allow for it in computing a reproduction rate would be to (a) compute

[12] Marital status is customarily disregarded in computing reproduction rates, which is equivalent to assuming that (a) the proportion ever married is the same at each age in the hypothetical cohort as in the actual population, or (b) differences in these proportions are of no consequence.

the probability of marrying for single women of each age in the base population, (b) apply these probabilities to the single women in the hypothetical cohort, (c) relate first births to ever married (rather than total) fecund zero parity women by age in the base population, and (d) apply the resulting probabilities to the ever married fecund zero parity women obtained in "b." This is impracticable because adequate information as to the number of first marriages by color and age of bride is lacking for most years.[11]

Under the circumstances the best procedure seems to be to (a) assume that all single women who bear a child marry before reaching the end of the childbearing period (or before dying), which probably is very close to the truth, and (b) allow for spinsterhood in the same way as for complete sterility, namely, by computing birth probabilities for, and applying them to, the women who may be expected to marry before they become too old to bear a child, which may well be set at age 50. Census reports show that approximately 10 per cent of the native white women aged 45–54 have been single in each census since 1890, the proportion varying from a high of 11.1 per cent in 1920 to a low of 8.2 per cent in 1890. In view of the narrow fluctuation of the proportion around 10 per cent this figure seems the best to apply to each cohort.

5. REPRODUCTION RATES WITH AND WITHOUT ADJUSTMENTS FOR PARITY, FECUNDITY, AND MARRIAGE

Net reproduction rates for native white women in the United States have been computed for each year from 1920 to 1944 from age, age-parity, and age-parity-fecundity-marriage specific birth rates. In two years the age adjusted rate is the same as the age-parity adjusted rate, being 105 in 1941 and 114 in 1944 in either case. (See Table 4.) In 18 years the adjustment for parity lowers the net reproduction rate by one or more points, the largest decreases being 11 in 1933 and 9 in 1932. In five years the adjustment raises the rate, the largest increases being 4 in 1921, and 3 in 1942 and 1943. Adjusting for either sterility or spinsterhood reduces the age-parity adjusted rate in 14 years, and raises it in 9 years, the maximum decrease being 7 in 1942 and the maximum increase 4 in 1933. Adjusting for parity, spinsterhood, and sterility gives results below the conventional age adjusted rates in every year except 1921 when it causes no change. The largest decreases—seven points—occur in 1931–33. Similar statements can be made for the gross reproduction rate and the intrinsic rate of natural increase.

Reproduction rates adjusted for age and parity, or for age, parity, sterility, and spinsterhood, show the wartime rise in fertility to have been somewhat larger than has been thought on the basis of the con-

[11] A complicating factor—the births that occur to single women—would necessitate the use of birth probabilities for single women as well as for married women.

ventional reproduction rates. From 1940 to 1943 the age adjusted net reproduction rate increased 21.0 per cent (from 100 to 121), the age-parity adjusted rate increased 27.6 per cent (from 98 to 125), and the

TABLE 4

GROSS AND NET REPRODUCTION RATES AND INTRINSIC RATES OF NATURAL INCREASE BASED ON AGE SPECIFIC, AGE-PARITY SPECIFIC, AND AGE-PARITY-FECUNDITY-MARRIAGE SPECIFIC BIRTH RATES FOR NATIVE WHITE WOMEN, UNITED STATES 1920–44

Year	Age Adjusted	Age-Parity-Adjusted	Age-Parity-Fecundity-Marriage-Adjusted	Year	Age Adjusted	Age-Parity-Adjusted	Age-Parity-Fecundity-Marriage-Adjusted
	*Gross Reproduction Rate**				*Intrinsic Rate of Natural Increase†*		
1944	122	122	110	1944	4.70	4.67	3.88
1943	130	134	128	1943	7.01	8.21	6.58
1942	125	128	120	1942	5.60	6.48	4.28
1941	113	113	110	1941	1.87	1.97	.85
1940	108	105	104	1940	−.03	−.84	−1.23
			Net Reproduction Rate*				
1944	114	114	111	1931	103	95	96
1943	121	125	119	1930	107	102	103
1942	116	119	112	1929	105	98	100
1941	105	105	102	1928	110	105	106
1940	99+	98	97	1927	116	115	114
1939	96	93	92	1926	118	115	115
1938	96	97	90	1925	123	122	121
1937	95	91	91	1924	127	128	126
1936	94	88	89	1923	125	123	122
1935	96	89	91	1922	124	123	122
1934	96	89	91	1921	133	137	133
1933	94	83	87	1920	125	127	124
1932	100	91	93				

* The rate per 100 persons; a stationary population has a rate of 100.
† The rate per 1,000 persons; a stationary population has a rate of zero.

age-parity-fecundity-marriage adjusted rate increased 22.7 per cent (from 97 to 119).

In interpreting the reproduction rates computed on the bases under consideration, it should be remembered that the proportion of women assumed to be completely sterile (10 per cent) was chosen as a maximum, hence that the reproduction rates based on age-parity-fecundity-marriage specific birth rates represent extreme values. The true values are between these and the rates adjusted for age and parity, but undoubtedly are closer to the former than the latter.

Although the age adjusted reproduction rates can be criticized from a methodological standpoint, the numerical differences between them and the age-parity adjusted rates or the age-parity-fecundity-marriage adjusted rates are not large. This is fortunate, of course, because the former are much simpler to compute than the latter. Furthermore, the

basic data required for age specific birth rates are available for a relatively large number of populations, whereas those required for age-parity, or age-parity-fecundity-marriage specific birth rates are available for relatively few populations.

Taking parity, sterility, and spinsterhood into account makes much larger changes in the prolificacy distribution of a hypothetical cohort than in its reproduction rates. The most striking difference occurs in the proportion of zero parity women in the hypothetical cohort of 1942. Childless women would constitute −8.4 per cent of the cohort[14] according to the age specific birth rates, 12.3 per cent according to age-parity specific birth rates, and 20.1 per cent according to age-marriage-fecundity-parity specific birth rates. But prolificacy distribution is a matter for discussion in another paper rather than here.

Even though reproduction rates based on age-parity-fecundity-marriage specific birth rates for women who marry and are not completely sterile will not be used widely in the near future the question of terminology may deserve some attention. It seems to the writer that two ideas should be kept in mind, namely, (a) the terminology used to date with reproduction rates has become well established and should be changed as little as possible, and (b) the fundamental concepts involved in the terms gross or net reproduction rate and intrinsic rate of natural increase have been refined rather than changed basically. One possibility would be to refer to the conventional rates as before, e.g., "net reproduction rates," but add the word "refined" in referring to the more accurate rates described in this paper, e.g., "refined net reproduction rates." The chief drawback is that "refined" has been used in the past in referring to rates adjusted for residence, age, and sex,[15] and, by itself, gives no clue as to the type of refinement. Other single words could only give a partial clue at best. It is suggested, therefore, that the phrases "age-adjusted" or "age-parity-fecundity-marriage adjusted" be used as prefixes. Thus the conventional net reproduction rate used heretofore would be called the "age-adjusted net reproduction rate," and the more accurate rates described here the "age-parity," or the "age-parity-fecundity-marriage adjusted net reproduction rate." Where space is at a premium, abbreviations could be used, e.g., "Age-Adj. Net Repro. Rate" and "Age-Par-Fec-Mar-Adj. Net Repro. Rate." It is realized that these terms are long, but it is believed that the disadvantage of their length is offset by the fact that they are relatively self-explanatory.

[14] There would be 1084 *first* births per 1000 women, hence 108.4 per cent of the women would have one or more births and −8.4 per cent would not bear a child.

[15] United States Bureau of the Census: "*Mortality Statistics*, 1925," Part II, Washington, G.P.O. 1920, p. 26.

THE PROBLEM OF NON-RESPONSE IN SAMPLE SURVEYS*

MORRIS H. HANSEN AND WILLIAM N. HURWITZ

The mail questionnaire is used in a number of surveys because of the economies involved. The principal objection to this method of collecting factual information is that it generally involves a large non-response rate, and an unknown bias is involved in any assumption that those responding are representative of the combined total of respondents and non-respondents.

Personal interviews generally elicit a substantially complete response, but the cost per schedule is, of course, considerably higher than it would be for the mail questionnaire method. The purpose of this paper is to indicate a technique which combines the advantages of both procedures.

The principle followed is to mail schedules in excess of the number expected to be returned, and to follow up by enumerating a sample of those that do not respond to the mail canvass. Under reasonable assumptions as to the relative costs of the two methods of canvass, an allocation of the sample can be made to mail and field canvasses. An illustration is given to show for a given degree of reliability, the varying sizes of the mailing list for different expected response rates, and the rate of field follow-up on the non-responses. For each response rate, the minimum cost of the survey is computed; from this computation it is possible to determine the maximum number of schedules to be mailed independent of the rate of response. Then to achieve the desired precision, the number to be interviewed would vary with the response rate actually found.

In a mathematical appendix the general formulas are derived.

THE MAIL questionnaire is used in a number of surveys because of the economies involved. The principal objection to this method of collecting factual information is that it generally involves a large non-response rate, and an unknown bias is involved in any assumption that those responding are representative of the combined total of respondents and non-respondents. Personal interviews generally elicit a substantially complete response, but the cost per schedule is, of course, considerably higher than it would be for the mail questionnaire method. The purpose of this paper is to indicate a technique which combines the advantages of both procedures.

The problem considered is to determine the number of mail questionnaires to be sent out and the number of personal interviews to take in

* This paper was presented at the annual meeting of the American Statistical Association on January 26, 1946 in Cleveland, Ohio.

following up non-responses to the mail questionnaire in order to attain the required precision at a minimum cost. The procedure outlined below[1] can be applied whatever the methods of collecting data are. For example, perhaps equally important as the problem of non-response in using mail questionnaires is the problem of call-backs in taking field interviews. In this latter problem the procedure to minimize cost for a given degree of reliability would call for taking a larger sample of first interviews and calling back on a fraction of "those not at home." The technique presented herein makes it possible to use unbiased designs at a reasonable cost where the excessive cost of ordinary methods of follow-up has frequently led to abandoning them.

To illustrate the principles, the simplest random sampling and estimating procedures are assumed in the early part of the text. The principles hold however for stratified sampling (see Section (b) page 525) and for other methods of estimation, as where a sample is used to estimate rate of change between two periods (see Section (c) page 526).

As an illustration, let us assume we want to estimate the number of employees in retail stores during a specified period in the State of Indiana. We shall assume we have a listing of all establishments having one or more employees, say from Social Security records, and their corresponding mailing addresses. A procedure sometimes followed is to take a sample of addresses from this list, mail out the questionnaires, and then depend exclusively on the mail returns for the estimate of number of employees for all retail stores in the State. The result of this procedure usually will be biased. It may be seriously so if there is a large rate of non-response. On the other hand, if all the addresses were actually visited by an enumerator, the cost of collecting the information would be much greater.

Suppose the cost of mailing is 10 cents per questionnaire mailed, and the cost of processing the returns is 40 cents per questionnaire returned. Suppose, on the other hand, that the cost of carrying through field interviews is $4.10 per questionnaire, and that this cost together with the cost of processing the field returns is $4.50 per questionnaire. For the cost of one field visit we could then obtain about eight mail questionnaires if there were a 60 per cent response and five mail questionnaires with only a 20 per cent response rate. This does not mean that we should take our entire sample by mail even though for the fixed cost we can make the actual sample size perhaps five to eight times as large as it would be if all the respondents were actually visited. It is

[1] The procedure given in this paper is an adaptation of the principles of double sampling developed by J. Neyman, see "Contributions to the theory of sampling human populations," *Journal of the American Statistical Association*, Vol. 33 (1938), pp. 101–116.

a common fallacy to assume that the size of the sampling error decreases in proportion to the square root of the number of schedules in the sample. Actually, in practice, this is seldom true. The sampling error depends on the over-all design of the sample of which the sample size is only one factor.

To illustrate, assume that questionnaires are sent to a sample of 1,000 addresses drawn at random from a complete list of 40,000 stores. Assume further that 500 or 50 per cent respond, and that of the 500 non-respondents 50 or 10 per cent are visited in order to insure some representation of the class of those that did not respond to the mail questionnaire. An unbiased estimate of the total number of employees is obtained from this sample by computing:

$$x' = \frac{N}{n}(m\bar{x}_1' + s\bar{x}_2''), \tag{1}$$

where

$N = 40{,}000 =$ the number of addresses on the mailing list;

$n = 1{,}000 =$ the number of questionnaires mailed out;

$\bar{x}_1' =$ the average number of employees per establishment for the stores responding to the mailed questionnaire;

$m = 500 =$ the number of such establishments;

$\bar{x}_2'' =$ the average number of employees per establishment for the field interviews;

$s = 500 =$ the number in the sample of the 1,000 questionnaires originally sent out that did not respond to the mailed questionnaire.

It is noted that the total size of sample actually processed would be equal to 500+50 or 550.

The sample variance of x' if the sampling method and estimating procedure specified are followed is given by:[2]

$$\sigma_{x'}^2 = N^2 \frac{N-n}{(N-1)n}\sigma^2 + \frac{N}{n}(k-1)\frac{S^2}{S-1}\sigma_b^2 \tag{2}$$

where

σ^2 is the variance in the entire population between the original establishments;

σ_b^2 is the variance among those not responding to the mailed questionnaire;

S is the number of establishments in the population that would not have responded to the mailed questionnaire had the mailed questionnaire been sent to all establishments;

[2] See appendix for development of this formula.

r is the number of establishments in the sample visited in the field;

and $k=\frac{s}{r}$, where, as already indicated, s is the number of non-respondents in the sample.

With this variance formula it can readily be shown that there are widely different sizes of samples which will have the same reliability and that a point will be reached where the size of sample alone will be a very poor indicator of the sampling reliability. For example, assume that $\sigma^2=\sigma_b^2$ and that N and S are sufficiently large that $\frac{N}{N-1}$ and $\frac{S}{S-1}$ are approximately equal to one. Further assume that the accuracy required is such that the average sampling error, ϵ, would be given by a sample of 1,000 when the rate of response is 100 per cent. If questionnaires were mailed to a random sample of $\hat{n}$ establishments and the response rate were 100 per cent, the variance of the total estimated from the sample would be given by

$$N^2\frac{N-\hat{n}}{(N-1)\hat{n}}\sigma^2$$

where σ^2 is defined as above. Thus

$$\epsilon^2 = N^2\frac{N-1{,}000}{(N-1)1{,}000}\sigma^2.$$

When we substitute various numerical values representing various proportions of mailed returns and field interviews for the symbols in formula (2) we find that there are a number of samples differing widely in size but each of which has the same average reliability.

Column 5 of Table 1 shows various sample sizes each of which yields the specified precision in the sample estimate. For example, sending out 1,500 questionnaires, obtaining a total of 1,125 questionnaires actually processed (750 by mail and 375 by field interview) yields exactly the same sampling error as sending out 10,000 questionnaires and obtaining a total of 5,263 questionnaires in the sample (5,000 by mail and 263 by field interview). It follows that at some point it would be unprofitable to put money into obtaining additional mail questionnaires and that it would be better to spend that money on obtaining returns from those not responding to the mail questionnaires.

Column 6 of Table 1 shows the total cost for each of the sample sizes

under the unit costs assumed in Table 1. Since the table is so constructed that the varying number of schedules tabulated all lead to exactly the same precision in the sample estimate, it is logical to pick that particular sample size that would lead to a minimum cost. The minimum cost will be achieved when 2,000 schedules are sent out, 1,000 of them are returned by mail (at the assumed 50 per cent response rate) and 333 of the non-respondents are interviewed in the field.

TABLE 1

SAMPLES OF DIFFERENT SIZES THAT LEAD TO SAME PRECISION OF RESULTS, THROUGH JOINT USE OF MAIL AND ENUMERATION METHODS ASSUMING A 50 PER CENT RESPONSE RATE

n	m	s	r	Schedules Tabulated	Cost
(1)	(2)	(3)	(4)	(5)	(6)
1,000	500	500	500	1,000	$2,550
1,500	750	750	375	1,125	2,138
2,000	1,000	1,000	333	1,333	2,009
2,500	1,250	1,250	313	1,563	2,159
3,000	1,500	1,500	300	1,800	2,250
4,000	2,000	2,000	286	2,286	2,487
5,000	2,500	2,500	278	2,778	2,751
10,000	5,000	5,000	263	5,263	4,184

C_1 = $0.10 = Cost per questionnaire of mailing
C_2 = $0.40 = Cost per questionnaire of processing returned questionnaires
C_3 = $4.50 = Cost per questionnaire of both enumerating and processing those obtained by field interviews
n = Number of questionnaires mailed out
m = Number of mail respondents
s = Number of non-respondents to mail canvass
r = Number of field interviews among the non-respondents

Instead of proceeding by trial and error as has been done above to determine the optimum numbers of schedules to mail out and to pick up by field interview, the optimum values of n and r can be computed from the following relatively simple formulas:

$$n = \hat{n}\{1 + (k - 1)Q\},$$
$$r = \frac{s}{k}, \tag{3}$$

where

$$k = \sqrt{\frac{C_3 P}{C_1 + C_2 P}}, \tag{4}$$

and

$$\hat{n} = \frac{N\sigma^2}{\sigma^2 + \epsilon^2 \frac{N-1}{N^2}}, \tag{5}$$

P is the rate of response to the mailed questionnaire, $Q = 1 - P$, and as indicated earlier, ϵ is the expected average error (standard error) to be tolerated in the total being estimated. Formulas (3) and (4) were obtained under the assumptions that $\sigma^2 = \sigma_b^2$, and $\frac{N}{N-1} \doteq \frac{S}{S-1} \doteq 1$. The optimum values of n and r without these assumptions are

$$\begin{aligned} n &= \hat{n}\left\{1 + (k-1)Q^2 \frac{N-1}{S-1} \frac{\sigma_b^2}{\sigma^2}\right\} \\ r &= \frac{s}{k}, \end{aligned} \tag{6}$$

with

$$k = \sqrt{\left\{\frac{N^2(S-1)\sigma^2}{S^2(N-1)\sigma_b^2} - 1\right\} \frac{C_3 Q}{C_1 + C_2 P}}. \tag{7}$$

Of course, the number of mail questionnaires and field interviews required to achieve a specified precision will vary with the response rate. In practice, one may not know even approximately what the response rate will be, whereas in order to estimate the optimum values from the above formulas, the approximate response rate must be known in advance of the survey. When the response rate is not known in advance one still may want to design the survey so as to achieve at least a certain specified precision at minimum cost, and at the same time to know about what the cost of taking the survey will be. Even under such circumstances it is possible to determine the optimum number of schedules to be sent out, and the optimum number of field interviews to be taken from among the non-respondents. For example, instead of assuming a 50 per cent response rate as in the above illustration, let us compute the optimum values of n and r for response rates varying from 10 per cent to 90 per cent, where we still want to achieve the same precision of results. The optimum values, computed from formulas (3) and (4), and the corresponding costs are shown in Table 2.

Column 4, Table 2 shows the cost of a survey with optimum joint use of mail and interview methods when the response rates are known.

Where the response rates are unknown, one alternative (referred to as Alternative 1 in the above table) which makes some use of the economies of using the mail questionnaire is to send out 1,000 questionnaires and follow up on all the non-responses regardless of the response rate. The sampling error would then be $N^2 \dfrac{N-1{,}000}{(N-1)1{,}000}\sigma^2=\epsilon^2$, no matter what the rate of response, and is the same as that for the other alternatives presented in Tables 1 and 2. However, for any given response rate,

TABLE 2
SAMPLES THAT LEAD TO SAME PRECISION OF RESULTS *THROUGH JOINT USE OF MAIL AND ENUMERATION* METHODS, FOR VARIOUS RESPONSE RATES

(Also comparison of the minimum costs for various response rates with the cost of sending out 1,000 questionnaires with a 100 per cent follow-up of the non-respondents.)

P	n	r	Cost of Optimum	Cost of Alternative 1
(1)	(2)	(3)	(4)	(5)
.10	1,714	900	$4,110	$4,190
.20	1,989	711	3,558	3,780
.30	2,034	575	3,035	3,370
.40	1,979	451	2,544	2,960
.50	1,870	341	2,006	2,550
.60	1,727	245	1,600	2,140
.70	1,564	163	1,328	1,730
.80	1,386	95	1,010	1,320
.90	1,107	40	731	910

the cost of this alternative will always be larger than the optimum. It is of interest to see how much more costly this procedure is than the optimum for various response rates. A comparison of columns (4) and (5) shows that the increase in cost over the optimum is smallest for the very low rates of response which is to be expected since not enough questionnaires have been received to take full advantage of the economies of the mail questionnaire. For response rates 30 per cent or greater, increases in cost of from 10 per cent to 24 per cent are to be expected if this alternative is used.

Where the approximate response rate is not known in advance, the following procedure, referred to as Alternative 2, is preferable to Alternative 1 described above because it generally makes more effective use of the economies possible through the mail questionnaire method. The first step is to determine the maximum number of questionnaires to be sent out no matter what the rate of response. The second step is

to determine the number to be interviewed in order to achieve the desired precision after the maximum number has been sent out and the rate of response is actually determined from the sample returns. Hence, the number to be interviewed would vary with the response rate actually found.

Note from column 2 of Table 2 that for the problem we are discussing, the maximum number to be sent out no matter what the rate of response, is 2,034 questionnaires. If 40 per cent responded, for example, then, using formula (3) we find that the number to be interviewed would be 448.

Table 3 shows, for varying rates of response to the mailed questionnaire, the number of field interviews to be taken to achieve the desired precision if 2,034 schedules are mailed, and the total cost for each of the response rates. The cost of the optimum that could have been used were the response rates known in advance is also shown.

TABLE 3

COMPARISON BETWEEN COST OF OPTIMUM IF RESPONSE RATE WERE KNOWN AND ALTERNATIVES 1 AND 2

p	m	r	Cost of Alternative 2	Cost of Alternative 1	Cost of Optimum
(1)	(2)	(3)	(4)	(5)	(6)
.10	203	862	$4,110	$4,190	$4,110
.20	407	710	3,561	3,780	3,558
.30	610	575	3,035	3,370	3,035
.40	814	448	2,545	2,960	2,544
.50	1,017	331	2,100	2,550	2,096
.60	1,220	227	1,713	2,140	1,690
.70	1,424	137	1,390	1,730	1,328
.80	1,627	66	1,161	1,320	1,010
.90	1,831	18	1,017	910	731

A comparison of column (4) with column (6) in Table 3 indicates that except for the very high response rates, the lack of any advance knowledge of the rates of response entailed almost no additional cost over the optimum values when the rates are known. When the rates of response are high, of course, the total cost of the survey will be small even though an unnecessarily large number of questionnaires had originally been sent out.

Thus, it can be seen that not only can optimum values of n and r be found for a response rate known in advance, but an optimum procedure can be found even where nothing is known in advance about the rate of response, and this procedure will produce results having at least the specified precision and at low cost. Of course, if the response

rate is known approximately in advance, the use of this information in determining the optimum use of mail and enumerative response will lead to slightly lower cost.

Some further comments on this problem:

(a) In actual practice a mail survey has a time limit. All schedules arriving before the deadline constitute the mail response and the field follow-up sampling ratio must be applied to all on the mailing list that did not respond before that date. The relatively few schedules arriving after that date, unless they are designated for interview, must be excluded from the sample, in order to avoid a bias of non-response of the type which we are trying to eliminate. The cut-off date of course should be held off until the mail response is substantially completed in order to take full advantage of the economies of the mail questionnaire. However, once a sample is designated for field follow-up and the respondent is actually interviewed in the field, the mail questionnaires returned (other than those designated for field follow-up) must be discarded.

(b) The optimum procedure described earlier for simple random sampling can be extended to stratified random sampling. Suppose, for example, that the population is divided into L strata with N_i establishments in each. If the costs per establishment do not differ widely between the different strata, a simple procedure is available for achieving economies through the joint use of the mail questionnaire method and field follow-up. The first step is to determine the size of sample required under the assumption of a 100 per cent response when allocation of the sample to the various strata is made in accordance with the well-known formula[3]

$$\hat{n}_i = \frac{N_i\sigma_i}{\sum N_i\sigma_i}\hat{n}, \tag{8}$$

where $\hat{n}$ is the size of sample required to achieve the required accuracy, ϵ. For stratified sampling $\hat{n}$ is approximately equal to $(\sum N_i\sigma_i)^2/\epsilon^2$.

The procedure is then to merely use formulas (3) or (6), for each stratum, to obtain the values for n_i, the number of mail questionnaires to be sent out in the i-th stratum, and r_i, the number of the non-respondents to the mail questionnaire to be taken for personal interview.

This procedure is not the optimum except under special conditions, but will be effective except in situations where the costs differ very

[3] J. Neyman, "On the two different aspects of the representative method; a method of stratified sampling and the method of purposive selection," *Journal of the Royal Statistical Society*, New Series, Vol. 97 (1934), pp. 558-606.

widely in the various strata. While the optimum value for r_i is the same as given in (6) when the subscript i is attached to each of the terms, the optimum value for n_i is more complicated than the corresponding value for the unstratified case. To determine the optimum values, in general, for stratified sampling, we first determine the optimum number of questionnaires to be sent out (n) and this is found to be equal to

$$n = \hat{n}\left\{\sum \frac{N_i^2}{\Phi_i}\frac{N_i}{N_i - 1}\sigma_i^2 + \sum \frac{N_i}{\Phi_i}(k_i - 1)\frac{S_i^2}{S_i - 1}\sigma_{bi}^2\right\}\frac{\sum \Phi_i}{\left(\sum N_i\sigma_i\sqrt{\frac{N_i}{N_i - 1}}\right)^2} \tag{9}$$

where

$$\Phi_i = \frac{k_i S_i \sigma_{bi}\sqrt{N_i}}{\sqrt{C_{3i}Q_i(S_i - 1)}}$$

and $\hat{n}$ is given by

$$\frac{\left(\sum N_i\sqrt{\frac{N_i}{N_i - 1}}\sigma_i\right)^2}{\left(\epsilon^2 + \sum N_i\frac{N_i}{N_i - 1}\sigma_i^2\right)}.$$

The optimum allocation of the n questionnaires to strata is then equal to

$$n_i = \frac{\Phi_i}{\sum \Phi_i}n. \tag{10}$$

(c) A ratio (or regression) type of estimate can be used instead of estimate (1) and at the same time make use of the optimum procedure. Thus if employment figures were available from a past census, an estimate of total employment which may be more efficient than estimate (1) consists of applying an estimate of the change in employment since the census date to the known number of employees at the census date. If we let

$\bar{y}_1'$ = Average number of employees at the census date, per establishment responding by mail;

$\bar{y}_2''$ = Average number of employees at the census date, per establishment in the field interview sample;

then an estimate of total employment would be

$$x'' = \frac{m\bar{x}_1' + s\bar{x}_2''}{m\bar{y}_1' + s\bar{y}_2''} Y \tag{11}$$

where

Y = Total employment at the census date, and m, s, $\bar{x}_1'$, and $\bar{x}_2''$ are defined as in estimate (1). An approximation to the variance of this estimate can be put into the same form as the variance $\sigma_{x'}^2$, except that the population variances that appear in the expression will now be the approximations to the variances of ratios. Then,

$$\sigma_{x''}^2 = N^2 \frac{N - n}{(N - 1)n} \sigma_1^2 + \frac{N}{n} \frac{S^2}{S - 1} (k - 1)\sigma_2^2 \tag{12}$$

where

$$\sigma_1^2 = \sigma_x^2 + R^2\sigma_y^2 - 2R\sigma_{xy}$$

and σ_x^2 is the same as σ^2 in formula (2); σ_y^2 is the corresponding variance of the number of employees at the census date; σ_{xy} is the covariance between x and y; and $R = \bar{x}/\bar{y}$ is the change in employment from the census date.

$$\sigma_2^2 = \sigma_{bx}^2 + R^2\sigma_{by}^2 - 2R\sigma_{b(xy)}$$

where σ_{bx}^2, σ_{by}^2, and $\sigma_{b(xy)}$ have the same meaning as the corresponding terms for σ_1^2 except that these refer to the non-respondent population.

Hence the optimum number of schedules to be mailed and field interviews to be taken can be determined as before merely by substituting σ_1^2 and σ_2^2 in formula (12) for σ^2 and σ_b^2 respectively in formula (2). The optimum formulas will then be given by (6) and (7).

APPENDIX

We shall now derive the variance of x' given by formula (2) and determine the optimum value of n and k, given by formulas (6) and (7).

$$\sigma_{x'}^2 = E(x' - x)^2 = E\left\{\frac{N}{n}(m\bar{x}_1' + s\bar{x}_2'') - x\right\}^2$$

$$= N^2E\left\{\frac{m\bar{x}_1' + s\bar{x}_2''}{n} - \bar{x}\right\}^2.$$

But $\bar{x}' = (m\bar{x}_1' + s\bar{x}_2')/n$ where $\bar{x}_2'$ = average for the s non-respondents. Therefore $m\bar{x}_1' = n\bar{x}' - s\bar{x}_2'$, and

$$\sigma_{x'}^2 = \frac{N^2}{n^2} E\{n(\bar{x}' - \bar{x}) + s(\bar{x}_2'' - \bar{x}_2')\}^2$$

$$= \frac{N^2}{n^2} \{En^2(\bar{x}' - \bar{x})^2 + 2En(\bar{x}' - \bar{x})s(\bar{x}_2'' - \bar{x}_2')$$

$$+ Es^2(\bar{x}_2'' - \bar{x}_2')^2\}. \quad (13)$$

Now for a fixed set of n observations, the quantities $\bar{x}'$, s, $\bar{x}$, and $\bar{x}_2'$ are also fixed, and $E\bar{x}_2'' = \bar{x}_2'$.

Therefore, the middle term of (13) vanishes. The third term is equal to

$$\frac{N^2}{n^2} s^2 \frac{s - r}{(s - 1)r} \frac{\sum_{i=1}^{s} (x_i - \bar{x}_2')^2}{s}.$$

Now the average or expected value over all subsets of exactly s non-responses is equal to

$$\frac{N^2}{n^2} s^2 \frac{s - r}{(s - 1)r} \frac{S}{s} \frac{s - 1}{S - 1} \sigma_b^2 = \frac{N^2}{n^2} s \frac{S}{S - 1} (k - 1)\sigma_b^2 \quad (14)$$

where σ_b^2 is the standard deviation between elements in the population of non-respondents, S is the number of non-respondents in the population and

$$k = \frac{s}{r}.$$

Since s varies from sample to sample, we must now take the expected value of (14) for all possible samples. This turns out to be

$$\frac{N^2}{n^2} \frac{S}{S - 1} (k - 1) \frac{nS}{N} \sigma_b^2 = \frac{N}{n} \frac{S^2}{S - 1} (k - 1)\sigma_b^2$$

and since

$$E(\bar{x}' - \bar{x})^2 = \frac{N - n}{(N - 1)n} \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N} = \frac{N - n}{(N - 1)n} \sigma^2,$$

$$\sigma_{x'}^2 = N^2 \frac{N - n}{(N - 1)n} \sigma^2 + \frac{N}{n} \frac{S^2}{S - 1} (k - 1)\sigma_b^2.$$

Let $\epsilon^2 = N^2 \dfrac{N - \hat{n}}{(N - 1)\hat{n}} \sigma^2$. We wish to find optimum values of n and k

such that the cost of the survey will be a minimum, for a fixed error, ϵ, in our estimate of x. The cost of the survey is expressible as

$$C = C_1 n + C_2 P n + C_3 \frac{Qn}{k}$$

where n is the number of questionnaires sent out and C_1 is the cost of mailing per questionnaire; Pn is the expected number of responses in the sample and C_2 is the cost per mail response; Qn/k is the expected number of field interviews and C_3 is the cost per field interview.

We then find the values of n and k which minimize the cost subject to the condition that $\sigma_{\bar{x}'}^2 = \epsilon^2$. These values turn out to be

$$k = \sqrt{\left\{\frac{N^2(S-1)\sigma^2}{S^2(N-1)\sigma_b^2} - 1\right\} \frac{C_3 Q}{C_1 + C_2 P}}$$

and

$$n = \hat{n}\left\{1 + (k-1)Q^2 \frac{N-1}{S-1} \frac{\sigma_b^2}{\sigma^2}\right\}.$$

CORRECTION

In the article, "Problems and Methods of the Sample Survey of Business" by Morris H. Hansen, William N. Hurwitz, and Margaret Gurney which appeared in the June, 1946 issue of this *Journal*, Formula 5, page 186, should read

$$n = \sqrt{\frac{C_1}{C_2}} \; \frac{\sigma_w}{\sqrt{\left(\sigma_b^2 - \frac{\sigma_w^2}{N-1}\right)}}.$$

ON THE ECONOMIC THEORY OF COST OF LIVING INDEX NUMBERS

MELVILLE J. ULMER
U. S. Department of Commerce

The concept of a "true" cost of living index has been carefully defined in theoretical economic-statistical literature, but the relation between this concept and officially published cost of living indices (ordinarily computed by Laspeyre's formula) has never been established. Theoreticians have shown only that $L > I_0$ and that $I_1 > P$, where: L is Laspeyre's index; I_0 is the true index based on the real income level of the base period; I_1 is the true index based on the real income level of the given period; P is Paasche's index.

Further analysis shows that definite theoretical relationships may be established among the differences, $L-I_0$, I_1-P, I_0-I_1, and $L-P$. These relationships make possible an estimate of the *maximum* difference likely to exist under ordinary circumstances between computed index numbers and theoretical indices, as follows:

$$L > I_0 > \frac{98.5}{100} L, \text{ and}$$

$$P < I_1 < \frac{101.5}{100} P$$

IDEALLY, a cost of living index ought to measure the change in money income required to yield equivalent satisfaction in two or more situations. This is the measure obviously required most frequently for practical purposes. It is also the definition generally adopted by theoreticians.[1] It is not, however, the definition of those who actually construct index numbers.

Unavoidably conscious of practical limitations, the makers of cost of living index numbers characteristically define their products purely in terms of the formula (usually Laspeyre's) used for their construction. The relation between these index numbers actually available for use, and the ideal, theoretically defined measure noted above has nevertheless remained persistently obscure. The users of these indices are incessantly tempted—wittingly or not—to infer a much broader application than the bare formula alone will allow. On the other hand, theo-

[1] For reference to the definitions employed by various writers in this field see Frisch, Ragnar, "Annual Survey of General Economic Theory: The Problem of Index Numbers," *Econometrica*, Jan. 1936, pp. 11–12.

reticians warn that Laspeyre's index is in no sense a reliable indicator of the true (theoretically defined) change in the cost of living.[3]

The present paper is an outgrowth of this dilemma. Its object is to demonstrate that under certain widely applicable conditions: (1) a

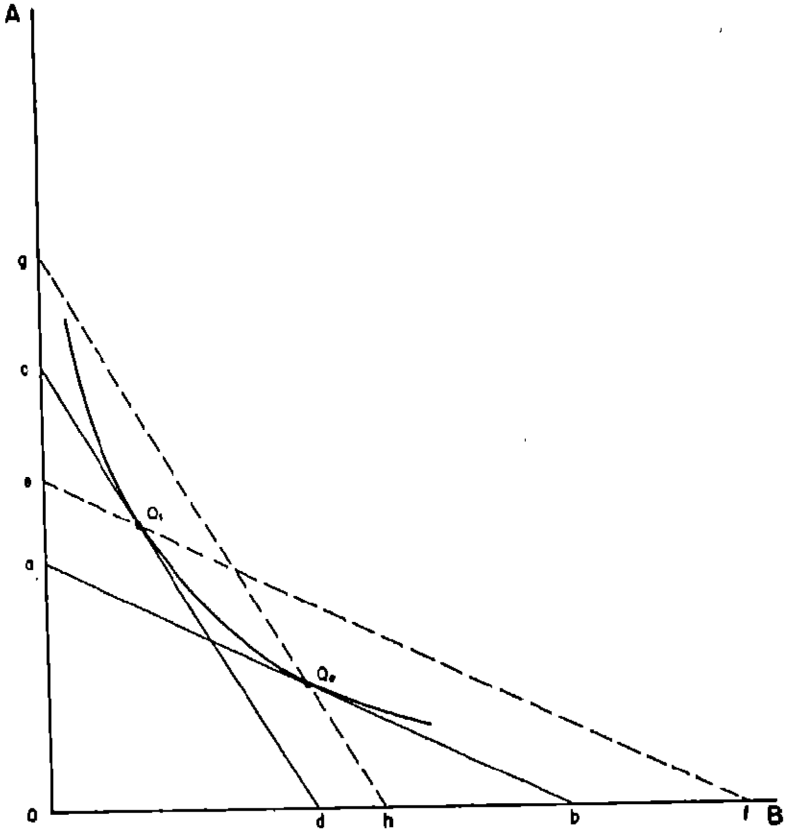


FIGURE 1

definite quantitative relationship may be established between Laspeyre's and the true index; (2) that from available evidence it may be concluded that Laspeyre's index is a close approximation of the true index; (3) that the maximum error of this approximation, likely to be encountered, may be determined. A knowledge of this relationship, and

[3] Cf. especially Staehle, H., "A Development of the Economic Theory of Price Index Numbers," *Review of Economic Studies*, June 1935, pp. 163–188; Frisch, Ragnar, *op. cit.*, pp. 1–38; Haberler, Gottfried, *Der Sinn der Index-Zahlen*, Tübingen, 1927; Mudgett, B. D., "The Cost of Living Index and Konüs' Condition," *Econometrica*, April 1945, pp. 171–181.

of the necessary conditions involved, should be of value not only to the users of cost of living index numbers but also to their makers.

THE RELATIONSHIP BETWEEN THE TRUE COST OF LIVING INDEX AND LASPEYRE'S FORMULA

Stated another way, the problem of measuring the true change in the cost of living consists: (1) of identifying equal real incomes in (say) two

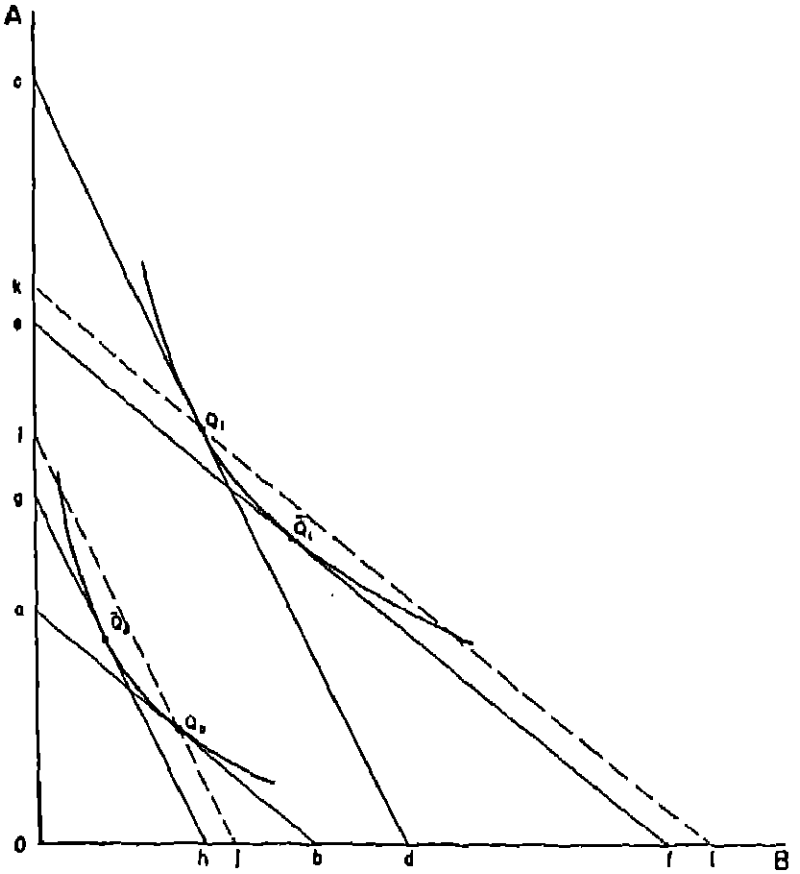


FIGURE 2

different situations; (2) of determining the ratio of the money values of these two real incomes. The result, then, would indeed be the theoretically "true" cost of living index as defined above. Strictly interpreted, a separate index of this kind is required for each distinguishable

real income level, though these need not necessarily differ numerically in all cases.

It has been shown, under certain highly restrictive assumptions, that Laspeyre's index and Paasche's index yield the upper and lower limits respectively of the true cost of living index. Although this demonstration is of no practical moment in itself—because of the narrowly limiting assumptions—it is a useful step in the derivation of more significant relations. It is described most conveniently in terms of indifference curves and expenditure lines.

In Figure 1 let the indifference curve shown represent a part of the indifference map of wage-earners for two commodities, A and B. By definition every point on this curve represents a bundle of goods of equivalent satisfaction. Let Q_0 and Q_1 represent two situations on this indifference curve. Let ab and cd represent expenditure lines tangent to the points Q_0 and Q_1 respectively. These lines are of the form $P_aA+P_bB=R$, where A and B are the two commodities, P_a and P_b are their respective prices and R equals total expenditures. The point of tangency *of an expenditure line to an indifference curve indicates the goods* which are purchased by consumers when they maximize their satisfaction under the given price system and the given total expenditures *indicated by the line.*

If total money expenditures in situation Q_1, in the accepted notation, equal Σp_1q_1 and in Q_0 equal Σp_0q_0 the ratio $\Sigma p_1q_1/\Sigma p_0q_0$ is the true cost of living index. Stated more fully, $\Sigma p_1q_1/\Sigma p_0q_0$ is the ratio of the total money expenditures required to yield equivalent satisfaction in the two situations—the definition of the true cost of living index.

In Figure 1 let commodity A be the *numéraire,* so that the intercepts of the expenditure lines on the A-axis will equal R, the total expenditures in each case. Then if we indicate the true cost of living index by the symbol I, we can write:

$$I = \frac{\Sigma p_1q_1}{\Sigma p_0q_0} = \frac{oc}{oa}.$$

Now in the same figure the expenditure line ef is drawn parallel to ab, and through the point Q_1. Since the the slope of an expenditure line *is given by the ratio* P_b/P_a, *the equality of slopes shows that* ab and ef are characterized by the same price system. Moreover, since ef passes through Q_1, its intercept, oe, indicates the total expenditures required to buy the commodities actually purchased in Q_1 under the price system prevailing in Q_0: that is, $oe=\Sigma p_0q_1$. In the same way, if gh is drawn parallel to cd, we can write $og=\Sigma p_1q_0$, representing total ex-

penditures required to buy the commodities actually purchased in Q_0, under the price conditions of Q_1.

From the figure, it will be noted that the following relationships hold,[3] all being necessary results of the fact that the normal form of an indifference curve is convex to the origin:

$$\frac{og}{oa} > \frac{oc}{oa}, \quad \text{since} \quad og > oc$$

$$\frac{oc}{oc} < \frac{oc}{oa}, \quad \text{since} \quad oc > oa.$$

If we substitute for these intercepts the expenditures they represent we obtain:

$$\frac{\sum p_1q_0}{\sum p_0q_0} > I > \frac{\sum p_1q_1}{\sum p_0q_1}.$$

In other words, Laspeyre's index is the upper limit and Paasche's index is the lower limit of the true cost of living index. This relationship, however, has been demonstrated only for a very special case. The two most important underlying assumptions are that tastes as well as real incomes[4] remain the same in the two situations compared.[5] Of course, the assumption that only two commodities appear in consumers' budgets is made here as well as elsewhere in this article for convenience only; it has been shown that the same relationship would hold for any number of commodities.

A more useful result is obtained if the most important of these assumptions is relaxed and we permit the situations to be compared to lie on different indifference curves. It is obvious, however, from the definition of a cost of living index, that if the situations to be compared are characterized by different real incomes, there must be *two* true indices (in the case of two situations), one based on the lower real income level and one based on the other. This is evident in Figure 2, where Q_0 and Q_1 are the two situations located on different indifference curves, while

[3] The following brief proof of the relationship $\Sigma p_1q_0/\Sigma p_0q_0 > I > \Sigma p_1q_1/\Sigma p_0q_1$ is the same as that given in H. Staehle, *op. cit.*, pp. 168–169.

[4] The sterility of this assumption for practical purposes was first pointed out by H. Staehle, *op. cit.*, p. 170. If it were actually known that real incomes were the same in the two situations compared, it would be unnecessary to compute Laspeyre's and Paasche's indices; the ratio of the actual money expenditures in the two situations, $\Sigma p_1q_1/\Sigma p_0q_0$, would give the true cost of living index exactly.

[5] It is curious to note that these assumptions are occasionally overlooked in practical work and the relationship $\Sigma p_1q_0/\Sigma p_0q_0 > I > \Sigma p_1q_1/\Sigma p_0q_1$ is explicitly taken as generally applicable. See, for example, Henry Shavell, "Price Deflators for Consumer Commodities and Capital Equipment, 1929–1942," *Survey of Current Business*, May 1943, p. 14.

ab and *cd* are the respective expenditure lines tangent to the indifference curves at these points. If *gh* is drawn parallel to *cd* and tangent to the indifference curve at point $\overline{Q}_0$, then point $\overline{Q}_0$ would represent the amounts of A and B which would be consumed at the lower utility level *at the prices prevailing in* Q_1. Hence, the ratio of the total expenditures in $\overline{Q}_0$ to those in Q_0 would indicate one true cost of living index; that is, the change in money income required to provide under the price conditions of Q_1, a *real* income equivalent in satisfaction to that received in Q_0. If commodity A is again taken as *numéraire*, then this true index may be represented by:

$$I_0 = \frac{\sum p_1\bar{q}_0}{\sum p_0q_0} = \frac{og}{oa}.$$

In the same way, by drawing the tangent *ef* parallel to *ab*, it can be shown that a second true cost of living index, based on the real income level which prevailed in Q_1, may be represented by:

$$I_1 = \frac{\sum p_1q_1}{\sum p_0\bar{q}_1} = \frac{oc}{oe}.$$

These two true cost of living indices are, of course, quite distinct, and it has been shown that they do *not* necessarily lie within the limits of Laspeyre's and Paasche's indices. Nevertheless, another relationship between these two formulas and the two true indices may be derived, the practical importance of which has been hitherto overlooked. If the expenditure line *ij* is drawn parallel to *cd* and through the point Q_0, then Laspeyre's index is represented by:

$$L = \frac{\sum p_1q_0}{\sum p_0q_0} = \frac{oi}{oa}.$$

Similarly, by drawing *kl* parallel to *ab* through point Q_1, we obtain Paasche's index:

$$P = \frac{\sum p_1q_1}{\sum p_0q_1} = \frac{oc}{ok}.$$

It is easily seen from the diagram that:

$$\frac{oi}{oa} > \frac{og}{oa} \quad \text{since} \quad oi > og, \quad \text{and}$$

$$\frac{oc}{oe} > \frac{oc}{ok} \quad \text{since} \quad ok > oe.$$

That is:

$$\frac{\sum p_1 q_0}{\sum p_0 q_0} > \frac{\sum p_1 \bar{q}_0}{\sum p_0 q_0} \quad \text{and} \quad \frac{\sum p_1 q_1}{\sum p_0 \bar{q}_1} > \frac{\sum p_1 q_1}{\sum p_0 q_1}, \quad \text{or} \tag{1}$$

$$L > I_0 \quad \text{and} \quad I_1 > P. \tag{2}$$

It is of course true, as other writers have contended, that these inequalities, *considered alone*, are of only very limited usefulness. They tell us only that I_0 is at some point below L, and I_1 is at some point above P; the relationships between L and P and between I_0 and I_1 are entirely unrestricted. Further analysis of these inequalities, however, permit much more useful conclusions, including a tentative estimate of the actual difference, to be expected in practice, between L and I_0, and between P and I_1, under certain conditions.

It is first of all important to note that I_0 and I_1, may differ for *one reason only*. They *may* differ only because real incomes have changed between the *two situations compared*; *actual* differences will *occur only* insofar as the change in real income results in altering the pattern of consumption between the two situations, relatively increasing (or decreasing) consumption of items which have advanced most in price, and relatively decreasing (or increasing) consumption of items which have advanced least. If the change in real income itself gave rise to *no* change in the pattern of consumption—i.e., if the elasticity of demand with respect to income were the same for *all* goods and services and equal to unity—I_0 would equal I_1. If there were no change in real income, of course, I_0 would equal I_1, in any case. Let the letter k represent the difference between I_0 and I_1.

On the other hand, differences between L and P may occur for two distinct reasons. The first is the same as that which accounts for any difference occurring between I_0 and I_1. If the difference k occurred between I_0 and I_1, then L and P would also differ by k, except insofar as this difference might be diminished or enhanced by the play of a second factor. Thus, if the elasticity of demand with respect to prices were zero for all commodities, then in the inequalities of (1) above, $\bar{q}_0$ would equal q_0 and $\bar{q}_1$ would equal q_1; hence L would equal I_0 and P would equal I_1. Differences between L and P and between I_0 and I_1 would therefore be the same, and in each case would be due to the possible effects of a change in real income upon the pattern of consumption. We have agreed to call this difference k.

The second factor making for differences between L and P is the possible alteration in patterns of consumption attributable to changes in

relative prices. It is that factor alone which accounts for the fact that L must be greater than I_0; the numerator of L is too large because it assumes that consumers do *not* alter their consumption in response to relative price changes, buying relatively more of cheaper items, and relatively less of more expensive items. For an analogous reason the denominator of P is too large, and P is therefore smaller than I_1. As already indicated, if the elasticity of demand with respect to prices were actually zero for all commodities, the difference between L and P would be k, the same as that between I_0 and I_1. Let us represent by the letter e the difference which occurs between L and P as the result of alterations in the pattern of consumption attributable to changes in relative prices. Then the *total* difference between L and P may be represented by:

$$L - P = [(L - I_0) + (I_1 - P)] + (I_0 - I_1) = e + k = d. \quad (3)$$

The quantity e, of course, is necessarily positive, Provided tastes remain unchanged (in accord with the assumption made earlier), its effect can be *only* to make L greater than P. The quantity k, however, may be positive or negative, and may be numerically greater or less than e. Hence d may be positive or negative.

Combining relationship (3) with the inequalities in (2), however, it is easily demonstrated that limits may be determined for the two true indices which should be of some practical importance. Since $I_0 - I_1 = k$ and $L - P = d$, we obtain from (2) by substitution:

$$\begin{aligned} & L > I_0 \quad \text{and} \quad I_0 - k > L - d \\ & L > I_0 > L - d + k \\ & L > I_0 > L - e. \end{aligned} \quad (4)$$

Similarly we obtain for I_1 the limits:

$$P < I_1 < P + e. \quad (5)$$

APPLICATION OF THE RELATIONSHIP

What does the relationship, $L > I_0 > L - e$, mean in quantitative terms? A precise answer to this question can be obtained only through direct measurement of e, a task clearly fraught with numerous statistical obstacles. For the interpretation of changes in the cost of living over time, as approximated by Laspeyre's index, however, an indirect approach is available which should prove satisfactory for most purposes.

The indirect approach is based on the fact that when k is positive, then d is positive and greater than c, by definition. Over a period of time in which the number of years characterized by a *higher* income level than that which prevailed in the base period, was roughly equal to the number of years characterized by a *lower* income level, it is probable that k would have a roughly equal chance of being positive or negative.[4] In any event, if the period of time were fairly long it is probable that k would be positive in a substantial number of cases. If the number of such cases were sufficiently large, then the highest value of d obtained during the period might be taken as an estimate of the highest value of d likely to be obtained in the future—an assumption which can be continuously tested with the passage of time. However, the highest value of d to be obtained can also serve as a conservative estimate of the maximum value likely to be obtained for c, since $d > c$ when $k > 0$.

Since d is easily measured this conclusion permits an estimate of the maximum error involved if L is taken as an approximation of I_0, or if P is taken as an approximation of I_1. Although measurements ideally suited to this purpose are not now available, the evidence which does exist, presented below, suggests that the value of d at its greatest is very small—probably in relative terms not more than $1\frac{1}{2}$ per cent. Hence, substituting this value in (4) and (5) above, we obtain the following limits to the two true indices:

$$L > I_0 > L - \frac{1.5}{100} L, \quad \text{or}$$

$$L > I_0 > \frac{98.5}{100} L, \quad \text{and} \tag{6}$$

$$P < I_1 < \frac{101.5}{100} P. \tag{7}$$

The existing evidence bearing on the value of d is shown in Tables 1 and 2. Although Table 1 does not actually present values of d, it has

[4] Whether k is positive or negative in any given case depends upon: (1) the direction of the change in real income from the base period, and (2) the elasticity of demand with respect to income for commodities whose prices are relatively flexible as compared with that for commodities whose prices are relatively inflexible. Thus, in the case of two commodities:

$$I_0 = \frac{\Sigma p_1 \bar{q}_0}{\Sigma p_0 q_0} = \frac{p_1' \bar{q}_0' + p_1'' \bar{q}_0''}{p_0' q_0' + p_0'' q_0''} \quad \text{and} \quad I_1 = \frac{\Sigma p_1 q_1}{\Sigma p_0 \bar{q}_1} = \frac{p_1' q_1' + p_1'' q_1''}{p_0' \bar{q}_1' + p_0'' \bar{q}_1''}$$

But $q' = f(p', p'', U)$ and $q'' = F(p'', p', U)$, where U is an index of real income level. Hence, substituting:

$$I_0 = \frac{p_1' f(p_1', p_1'', U_0) + p_1'' F(p_1'', p_1', U_0)}{p_0' f(p_0', p_0'', U_0) + p_0'' F(p_0'', p_0', U_0)} \quad \text{and} \quad I_1 = \frac{p_1' f(p_1', p_1'', U_1) + p_1'' F(p_1'', p_1', U_1)}{p_0' f(p_0', p_0'', U_1) + p_0'' F(p_0'', p_0', U_1)}$$

supplementary interest. In this table the two measures shown are both fixed base (Laspeyre's) indices. One of these measures of changes in the cost of living, however, is based on average purchases of wage-earners in 1917–1918; the other is based upon purchases in 1934–1936. Most notable is the fact that the percentage difference between these indices during the period covered is at most 1.1 per cent. The clear implica-

TABLE 1

PER CENT DIFFERENCE BETWEEN OLD AND CURRENT COST OF LIVING INDICES OF THE BUREAU OF LABOR STATISTICS

Date		Indices 1935–39 = 100		Per cent Difference,* Old to Current
		Old Index	Current Index	
		Fixed budget priced: 1917–1918 average purchases	Fixed budget priced: 1934–1936 average purchases	
1935	Mar.	97.8	97.8	0
	July	97.6	97.6	0
	Oct.	98.0	98.0	0
1936	Jan.	98.7	98.8	.1
	Apr.	97.9	97.8	.1
	July	99.6	99.4	.2
	Sept.	100.0	100.4	.4
	Dec.	100.0	99.8	.2
1937	Mar.	101.7	101.8	.1
	June	102.6	102.8	.2
	Sept.	103.2	104.3	1.1
	Dec.	102.6	103.0	.4
1938	Mar.	100.7	100.9	.2
	June	101.2	100.9	.3
	Sept.	100.4	100.7	.3
	Dec.	100.4	100.2	.2
1939	Mar.	99.6	99.1	.5
	June	99.2	98.6	.6
	Sept.	100.4	100.6	.2
	Dec.	99.8	99.6	.2

* Signs ignored.
Source: U. S. Bureau of Labor Statistics.

tion is that even over long periods of time—in this case 17 years—stability in the pattern of consumption may be very great.

Table 2 presents values of d from 1929 to 1940, but not of the kind ideally required for the purpose set forth in this paper. Ideally, d should be computed from indices measuring changes in the cost of living for some fairly homogeneous group sharing a roughly common standard of living; the cost of living index in which chief interest lies, of course, is that for urban wage-earners. The retail price indices of

Table 2, however, measure price changes only for sales in retail stores, and of course cover *total* sales of this kind in the United States.

Nevertheless, in the absence of better data, the results are of some interest. They show notably low values of d. Clearly, the size of d is subject to cyclical swings, reaching maximum values regularly at the peaks and troughs of business cycles.[7] Indeed, taken in conjunction with Table 1, there is the suggestion that cyclical swings in real incomes and relative prices may have greater effects upon the pattern of consumption than the secular changes over a period as long as 17 or

TABLE 2

VALUES OF d, BASED ON RETAIL PRICE INDICES OF DEPARTMENT OF COMMERCE

Date*	Indices 1939=100		d
	L	P	
	Fixed Weighted Index; Weights based on Sales in 1939	Variable Weighted Index; Weights based on Sales in given year	Expressed as Per cent of L†
1929	131.8	129.9	1.4
1930	124.3	123.6	.6
1931	107.0	106.9	.1
1932	92.7	92.4	.3
1933	90.4	90.2	.2
1934	96.6	96.7	.1
1935	102.3	102.0	.3
1936	102.4	102.1	.3
1937	106.3	105.9	.4
1938	101.7	101.7	.0
1939	100.0	100.0	—
1940	101.2	101.1	.1

* Index numbers for war years subsequent to 1940 are not suited to the purpose of this table and therefore are omitted; see section below on "Some Economic Implications."

† The values of d, when expressed as percentages of P are the same for every year except 1929. In 1929 the value of d, when expressed as a per cent of L, is 1.44; when expressed as a per cent of P, it is 1.46.

Source: U. S. Department of Commerce.

18 years. Nevertheless, it is most important to note that the maximum value of d attained during the entire period is only 1.4 per cent. It is on the basis of this maximum value—rounding the figure to 1½ per cent[8]—that the limits to the true indices in (6) and (7) were established.

[7] Cf. the remarks of Frederick C. Mills in Black, John D. and Mudgett, Bruce D., *Research in Agricultural Index Numbers*, Social Science Research Council, *Bulletin* No. 10, 1938, p. 83.

[8] The value of d in 1929 when stated in relative terms with Laspeyre's index as base is, carried to two places, 1.44 per cent; when stated relatively in terms of Paasche's index, the value of d is 1.46 per cent.

Thus, applying these limits, if Laspeyre's index in any given year were 110, then the true index (I_0), based on the base period real income level, would lie between 110 and 108.4. Similarly, if Paasche's index stood at 109, then the true index (I_1), based on the given year's real income level, would lie between 109 and 110.6.

This analysis shows, therefore, that Laspeyre's index, may indeed be considered an approximation of the corresponding true index, and that the maximum error of this approximation may be estimated. An analogous statement applies to Paasche's index. Moreover, although the data above by no means constitute conclusive evidence,[9] there is a clear indication that the error involved in these approximations is most likely very small.

SOME ECONOMIC IMPLICATIONS

In conclusion, a word concerning the economic implications of these findings may be in order. The principal assumption upon which the above analysis was based, as previously noted, is that tastes do not change appreciably over the situations compared. There was also the implicit assumption, obviously closely related, that over the time period considered the bulk of goods and services remain about the same in variety and quality. These assumptions underlie all theoretical analyses of cost of living index numbers, and it appears agreed that—except during periods of violent social upheaval such as war—they offer sufficiently closely approximations of reality provided the period considered is not too extended.[10]

Now it has been indicated above that, theoretically, the Laspeyre and Paasche index numbers would exactly equal the corresponding true cost of living indices if (1) there were no changes in *relative* prices from time to time, or if (2) the elasticity of demand with respect to

[9] The writer has been informed by officials of the U. S. Bureau of Labor Statistics that the Bureau plans to conduct periodical studies of wage-earners' budgets—probably once a year. These data, when available, will make possible the construction of Paasche's index periodically, and, in turn, measurements of d suited directly to the use proposed in this article.

[10] Cf., for example, Staehle, *op. cit.*, p. 104. Stability in the pattern of consumption over time for a clearly defined group such as wage-earners, as well as for the nation as a whole, has been one of the outstanding results of budget studies conducted in this country. This and other evidence demonstrate beyond question that changes in tastes are usually gradual, although accelerated by cyclical influences, and ordinarily affect only a few commodities even over fairly long periods. Similarly, the variety and quality of the bulk of all goods offered for sale ordinarily remain about the same over reasonably extended periods. The fact that minor changes *do* occur rather frequently can in some measure be offset if the maker of the index number concerned puts to use (so far as possible) what Keynes termed the "method of equivalent substitution." (See *A Treatise on Money*, Vol. I, 1930, pp. 103–104.) Of course, if it is believed that fundamental changes have occurred in either the tastes of consumers or the kinds of goods offered for sale, or both, neither Laspeyre's nor any other index can be said to approximate the true index; from a practitioner's point of view, the continuity of the index must under these circumstances be interrupted in order to change the composition of the base period budget priced.

prices were zero for all goods and services. The first condition means that all prices change from time to time by equal relative amounts. The second condition means that goods and services are consumed in constant proportions regardless of changes in relative prices (although proportions may change in response to variations in money prices and incomes). Obviously, neither of these conditions is tenable.

However, it has been contended in this paper that the Laspeyre and Paasche index numbers are *close approximations* to (though they are not identical with) the true cost of living indices. This raises a question, with respect to the above conditions, of "more or less" rather than of "either or." The aggregate amount of expenditures wage-earners are required to make in order to maintain a given plane of living is affected (A) by changes in money prices. It is also affected by (B) changes in relative prices and the adjustments in consumption patterns made by consumers *in response to these relative price changes.* The effects of factor (A) are measured by the true cost of living indices as well as by the Laspeyre and Paasche index numbers. The effects of factor (B) are measured only by the true indices. In economic terms, therefore, the findings of this paper suggest that the effects of factor (B) are ordinarily very small, perhaps most often negligible, when compared with the effects of factor (A).[11]

[11] An analysis of price behavior and of consumer expenditures from this point of view should probably bear this out. Most analyses are designed to highlight *differences* among commodity price and consumption changes, for it is these differences which provide insight into economic behavior. Actually, however, the broad similarities which persist over time far outweigh the differences. While relative prices, in detail, change constantly, the *general order of relative prices remains always about the same.* For example, to the writer's knowledge tenderloin steak is at all times priced considerably higher than beef liver, even though the exact differential may and does change. Similarly, even though consumers change their pattern of consumption in response to prices, the persistency of common needs, habits and conventions is obviously far more influential than the occasional modifications.

THE COMPUTATION OF PARTIAL CORRELATION COEFFICIENTS

FREDERICK V. WAUGH
Office of War Mobilization and Reconversion

If we have already computed the partial regression of x_1 on x_k, $b_{1k\cdot 23\ldots)k(\ldots n}$, and its standard error, $Sb_{1k\cdot 23\ldots)k(\ldots n}$, it is a simple matter to compute the partial correlation between x_1 and x_k, $r_{1k\cdot 23\ldots)k(\ldots n}$. To simplify the notation we shall designate the partial regression as $b_{1k\cdot}$, its standard error by $Sb_{1k\cdot}$, and the partial correlation as $r_{1k\cdot}$.

A simple relation between these measures is

$$r_{1k\cdot} = \frac{b_{1k\cdot}}{+\sqrt{b_{1k\cdot}^2 + N'S^2b_{1k\cdot}}} \tag{1}$$

where N' represents the number of degrees of freedom,—that is, N' equals the number of observations, N, minus the number of variables, n. If we want the corrected partial correlation coefficient, $\bar{r}_{1k\cdot}$, we simply replace N' by N in the denominator of (1).

So far as I know this formula for computing partial correlations is new. Certainly the computational effort is much less than that required by methods in common use. They are not well adapted to the computation of partial correlation coefficients.

We shall illustrate the use of (1) by reference to two examples chosen from Ezekiel's well-known text.[1] On pages 469–477 Dr. Ezekiel shows in detail how the so-called "Doolittle Method" can be used to compute various measures of correlation and regression in the case of a given numerical problem of four variables. Using x_1 as the dependent variable, he computes the following regressions with their respective standard errors:

$$b_{12\cdot 34} = -0.810 \pm 0.233$$
$$b_{13\cdot 24} = 0.180 \pm 0.030$$
$$b_{14\cdot 23} = -0.390 \pm 0.159$$

His computation of the three partial correlations requires three separate, and additional, applications of the Doolittle Method. If we use the method suggested in this paper this work is unnecessary. Using (1), with $N=13$ and $N'=13-4=9$, we get

[1] Mordecai Ezekiel, *Methods of Correlation Analysis*, second edition, 1941.

$$r_{12\cdot 34} = \frac{-0.810}{+\sqrt{(0.810)^2 + 9(0.233)^2}} = -0.757$$

$$r_{13\cdot 24} = \frac{0.180}{+\sqrt{(0.180)^2 + 9(0.030)^2}} = 0.896$$

$$r_{14\cdot 23} = \frac{-0.390}{+\sqrt{(0.390)^2 + 9(0.159)^2}} = -0.544.$$

The above simple computations replace Ezekiel's Tables 93 and 94 and the computations on page 477.[2]

In the other illustration we shall compute the absolute values of several partial correlations when we know only the ratio, $Sb_{1k\cdot}/b_{1k\cdot}$. Dividing both the numerator and denominator of (1) by $b_{1k\cdot}$, we have

$$|r_{1k\cdot}| = 1/\sqrt{1 + N'S^2b_{1k\cdot}/b_{1k\cdot}^2}. \qquad (2)$$

This equation is useful because occasionally only the ratios are given rather than the coefficients themselves.

A case in point is a table in Ezekiel, summarizing the results of two regression analyses based upon two studies of factors related to milk production per cow.[3] The first two columns of Table I show the standard error ratios as given. From these we compute the last two columns, using equation (2), with $N' = 85$ in the case of the Wisconsin study, and with $N' = 65$ in the case of the Minnesota study.

To determine the signs of the partial correlations in this table we would need to know only the signs of the partial regressions, since the sign of any partial correlation coefficient is the same as the sign of the corresponding partial regression coefficient.

The sixteen partial correlations in Table 1 were computed in slightly less than twenty minutes. It would have taken many hours to compute them by sixteen applications of the Doolittle Method.

Derivation of formulas. Formulas (1) and (2) can be derived from two previous papers by the present writer.[4]

[2] Incidentally, the partial correlations computed here do not agree with those in the first printing of Dr. Ezekiel's second edition. A new printing of this edition has just been made, giving the correct partials, which check with those given here.

[3] *Op. cit.*, page 327, Table 75. We have converted the probable error ratios to standard error ratios by multiplying each by 1.48258.

[4] F. V. Waugh, "A Simplified Method of Determining Multiple Regression Constants," *Journal of the American Statistical Association*, Vol. 30 (1935), pp. 694–700.

F. V. Waugh, "A Note Concerning Hotelling's Method of Inverting a Partitioned Matrix," *Annals of Mathematical Statistics*, Vol. 16 (1945), pp. 216–217.

TABLE 1

COMPUTATION OF ABSOLUTE VALUES OF PARTIAL CORRELATIONS BETWEEN MILK PRODUCTION PER COW AND SEVERAL OTHER VARIABLES

	Relative error		Absolute Values of Partial Correlations	
	Wisconsin study	*Minnesota* study	*Wisconsin* study	*Minnesota* study
Total digestible nutrients	.178	.170	.520	.589
Nutritive ratio	.184	.141	.508	.600
Per cent of protein "good"	.420		.250	
Per cent of lime	.000		.108	
Per cent of summer feeding	.269		.386	
Per cent silage	.310	.203	.322	.521
Fat test of milk	.157	.055	.508	.014
Per cent fall freshening	.274	.175	.308	.579
Value per cow	.397		.310	
Age of cows		.205		.424
Per cent grain in ratio		.290		.383

Briefly, the first of these papers showed that if the inverse of the complete moment matrix is D, and its elements are d_{ij}, the partial correlation coefficient

$$r_{1k\cdot} = \frac{-\,d_{1k}}{+\sqrt{d_{11}d_{kk}}}; \tag{3}$$

that

$$d_{11} = 1/(m_{11} - b_{12\cdot}m_{12} - b_{13\cdot}m_{13} - \cdots - b_{1n\cdot}m_{1n}); \tag{4}$$

and that

$$d_{1k} = -\,d_{11}b_{1k\cdot}. \tag{5}$$

Thus, after the regression coefficients have been computed (4) and (5) give us two of the three values we need in (3). The missing value is that of d_{kk}. This value, as well as that of d_{11} and d_{1k} are computed by the procedure outlined in my 1935 paper. But the methods explained by Fisher and Ezekiel are based upon computing the inverse of the *incomplete* moment matrix, obtained by eliminating the row and column representing the dependent variable. Fisher and Ezekiel, thus, compute the inverse matrix, C, the elements of which are c_{ij}.

But, when we have computed c_{ij}, d_{11}, and $b_{1k\cdot}$ we can use the results of my 1945 paper to show that[5]

$$d_{ij} = c_{ij} + d_{11}b_{1i\cdot}b_{1j\cdot}. \tag{6}$$

[5] Both my paper and Harold Hotelling's, "Some New Methods of Matrix Calculation," *Annals of Mathematical Statistics*, Vol. 14 (1943), pp. 1–34, were presented as means of computing the inverse of a square matrix of $2p$ rows that was partitioned into four p-rowed matrices. But the same results can be obtained if any square matrix is so partitioned that the matrix labeled, a, in my paper is square.

It is possible to use (6) to compute the complete inverse, D. Formulas (1) and (2) require only the computation of the diagonal elements, d_{kk}. These can be written

$$d_{kk} = c_{kk} + d_{11}b^2_{1k\cdot}. \tag{7}$$

Inserting in (3), the values in (5) and (6), we have

$$r_{1k\cdot} = \frac{d_{11}b_{1k\cdot}}{+\sqrt{d_{11}(c_{kk} + d_{11}b^2_{1k\cdot})}} = \frac{b_{1k\cdot}}{+\sqrt{\dfrac{c_{kk}}{d_{11}} + b^2_{1k\cdot}}} \tag{8}$$

But, since

$$S^2_{b_{1k\cdot}} = \frac{c_{kk}}{N'd_{11}},$$

(8) can be written

$$r_{1k\cdot} = \frac{b_{1k\cdot}}{+\sqrt{N'S^2_{b_{1k\cdot}} + b^2_{1k\cdot}}}.$$

Since writing this paper I have found that Dr. M. A. Girshick of the United States Department of Agriculture has worked independently upon the problem of increasing the order of an inverse matrix, and has actually computed the complete inverses of a number of matrices by the repeated use of (6). Dr. Girshick and I have developed simple work sheets for using this method to compute successively a 1-rowed, 2-rowed, . . . , *n-rowed matrix, by adding a row at each step.* The number of multiplications appears to be almost identical to the number required by the Doolittle Method.

AN APPLICATION OF SEQUENTIAL S[illegible] TESTING STUDENTS

DUDLEY J. COWDEN
University of North Carolina

This paper provides an illustration of an application to the field of education of a technique hitherto used mainly in industry. The effect of systematic stratified sampling is briefly discussed.

STATISTICAL methods have recently been developed for the control of quality in industry which are adaptable to the field of education. In industry quality control serves two purposes: (1) control of the process of manufacture; (2) control of the quality of accepted lots. In acceptance control one or more samples from each manufactured lot is inspected to see whether the lot should be accepted or rejected. It would be desirable to accept all good lots and to reject all bad lots; but when dealing with samples one cannot avoid making an occasional error of rejecting a good lot or of accepting a bad lot. Having defined what is meant by a good lot (say one containing one-half per cent or less of defective items) and a bad lot (say one containing five per cent or more of defective items) it is then possible to set standards of acceptance and rejection which will give any desired degree of assurance against making either of the above mentioned types of error. If one is to select a single sample, however, and base his decision on that sample, the size of the sample required may be very large, and the amount of inspection, therefore, very large. Furthermore, for some products the testing is destructive, and it is highly desirable to have the sample size as small as is consistent with the requisite ends. Consequently, a very interesting procedure has been worked out, known as *sequential sampling*,[1] whereby items are tested one at a time (or in small groups) and a decision to accept or to reject the lot made as soon as enough data have been accumulated to justify one of these decisions. One never knows in advance how many items will be required to reach a decision, although an estimate can be made of the average size of sample that will be needed. If the quality of the lot is very good or very bad a decision

[1] The theory of sequential sampling was worked out by Abraham Wald. For the mathematical theory see Abraham Wald, "Sequential Tests of Statistical Hypotheses," *The Annals of Mathematical Statistics*, Vol. XVI, June 1945, pp. 117–186. For a shorter and less mathematical treatment see Abraham Wald, "Sequential Method of Sampling for Deciding between Two Courses of Action," *Journal of the American Statistical Association*, Vol. 40, September 1945, pp. 277–306. For detailed instruction concerning the use of the method see Statistical Research Group, Columbia University, *Sequential Analysis of Statistical Data: Applications*, Columbia University Press, 1945.

will generally be reached with a small sample; if the quality of the lot is of an intermediate character a very large sample may be required.

When an examination is given to a student it sometimes happens that not enough questions are asked to permit a fair evaluation of his knowledge and ability. On the other hand the examination is sometimes drawn out longer than is necessary. If a student is very good or very poor only a few questions may be needed to establish this fact beyond reasonable doubt; but borderline students should be examined at considerable length before deciding whether they should be passed or failed. If sequential sampling is used the fate of good students and of poor students tends to be determined quickly, but mediocre students must continue with the examination until the results give adequate grounds for a decision. By use of the sequential method the number of questions answered by a student is reduced to a minimum, and at the same time the probability of passing a poor student or failing a good student is controlled.

The writer recently conducted an experiment with the method of sequential analysis as part of a final examination in a small class in elementary statistics at the University of North Carolina. The method was used in administering a true-false examination. This examination consists of 200 questions, classified into ten sections according to subject matter with a varying number of question in the different sections and arranged within classes according to degree of difficulty. These 200 questions may be regarded as a sample from a much larger list of similar questions which could be asked. Past experience over several years indicates that the test scores should be graded as in Table 1.

TABLE 1

PER CENT OF ERRORS AND GRADES ON A TRUE-FALSE EXAMINATION

Per cent of errors	Grade
20 or less	A
More than 20 but including 27	B
More than 27 but including 32	C
More than 32 but including 35	D
More than 35 but including 38	E
More than 38	F

On the basis of a particular set of questions it would be ideal to pass every student who could answer more than 65 per cent of all questions of the type included in this examination, and fail every student who could answer less than 65 per cent. But of course a D-grade student might be so unfortunate as to draw a set of questions that was

unusually difficult for him, while an E-grade student might be lucky enough to pass because he drew a number of questions that were quite to his liking. Consequently we must make some compromises. First, we must decide the proportion of good students we are willing to fail and the proportion of poor students we are willing to pass. Also (unless these two proportions are to be complementary, such as .8 and .2, or .4 and .6) we cannot make a hairline distinction between a good student and a poor student, but must leave a zone of mediocrity into which some students may fall. That this is true may be seen for the case of students who are able (in the long run) to pass exactly 65 per cent of all possible questions similar to those in this examination. If we fail 20 per cent of these students on a given test we must of course pass the other 80 per cent. But if we arbitrarily decide that a good student is one who can answer 70 per cent of the questions, while a poor student is one who can answer only 60 per cent of the questions, the proportion of good students whom we fail need bear no relationship to the proportion of poor students whom we pass. The finer the distinction we make between a good student and a poor student the larger the number of questions we must ask in order to have the desired assurance that the right decision will be made.

In the present instance it was decided, largely on the basis of judgment, that we were willing to fail not more than 20 per cent of all students who could answer 70 per cent or more of the questions, while we were willing to pass not more than 10 per cent of all students who could answer 60 per cent or less of the questions. Thus we were willing to fail one out of every five C-grade students and to pass one out of every ten F-grade students. Let us adopt these conventional symbols:

p_1, the maximum proportion of errors in all possible questions of a given type made by a student who is definitely good;

p_2, the minimum proportion of errors in all possible questions of a given type made by a student who is definitely poor;

α, the probability of failing a good student;

β, the probability of passing a poor student.

The numerical values which we have determined, mainly subjectively, for these constants are:

$$p_1 = .30 \qquad \alpha = .20$$

$$p_2 = .40 \qquad \beta = .10$$

It may seem outrageous to the reader that we should be willing to fail one of every five C-grade students. In this case, however, an injustice was not inflicted, since at the end of each "round" of 20 questions the

student was given the option (1) of accepting his grade, providing the results were determinate, or (2) of continuing with the examination. It was more important not to pass very poor students, since such students would not usually exercise their option of continuing with an examination which they had already passed. Also it must be remembered that the true-false questions constituted only half of the final examination; the other half consisted of problems. Finally, the grade in the course was only partly determined by the grade on the final examination; five one-hour quizzes were given during the term.[3] With respect to the true-false test, if the results were indeterminate at the end of 20 questions the student was required to take 20 more, making 40 altogether. If the results were still indeterminate, 20 more were answered, for a total of 60 questions. No student was required to continue beyond 60 questions, nor permitted to continue beyond 100 questions. Some upper limit was necessary because the period of time allotted to the examination was limited.

Using D_1 *(decision number 1) to indicate the number of questions* that can be missed and still permit a student to pass, D_2 (decision number 2) to indicate the number of questions that must be answered wrong before a student is *failed*, and N to indicate the cumulative number of questions answered, the method of sequential analysis has been worked out in such a way that two linear equations result:

$$D_1 = a_1 + bN;$$
$$D_2 = a_2 + bN.$$

As can be seen, the straight lines representing these two equations are parallel, and differ only as to the constants a_1 and a_2. These constants depend on the values of p_1, p_2, α, and β adopted. The more widely p_1 and p_2 differ the closer together the two lines will be, and therefore the more quickly will a decision be reached. The larger the value of α or β the smaller will be the value of a_2, and the larger (algebraically) will be the value of a_1. Therefore to bring the two lines closer together we must increase α and/or β. a_1 is always negative, since answering all questions correctly does not very strongly indicate knowledge of the subject until a reasonable number of questions are answered (what is a reasonable number depends on the value adopted

[3] We thus have seven observations on the ability of each student. In case the average or the range of grades for any student is too far out of line in comparison with the rest of the class, this student may be considered "out of control"; and his case may be further considered (as by being given a condition in the course). In industry a "control chart" is generally used for a similar purpose. For a simple exposition of the control chart method see American Standards Association, *Control Chart Method of Controlling Quality During Production (ASA Z1.3–1942)*, New York, 1942.

for β, becoming larger as β is made smaller). On the other hand, a_2 is always positive; but a decision to fail cannot be reached until $D_2 \leqq N$, since a student cannot miss more questions than he answers. when $\alpha=\beta$, $a_2=-a_1$. The slope b is independent of α and β, but depends exclusively on p_1 and p_2. b is not only the slope of the two lines, but it is the limiting value for per cent of errors for D_1 and D_2 as N approaches infinity (see Columns 3 and 6 of Table 2). In the present instance the values (to two digits) of p_1 and p_2 were chosen so as to make the value of b, which is the dividing line between passing and failing, .35 when rounded to two digits.

The constants a_1, a_2, and b are easily computed after certain intermediate values g_1, g_2 h_1 and h_2 have been obtained. These intermediate values are also useful for other purposes, which will be explained. The formulas are given below, and also their application to the present illustration:

$$g_1 = \log \frac{p_2}{p_1} = \log \frac{.4}{.3} = .124938;$$

$$g_2 = \log \frac{1-p_1}{1-p_2} = \log \frac{.7}{.6} = .066947;$$

$$h_1 = \frac{\log \dfrac{1-\alpha}{\beta}}{g_1+g_2} = \frac{\log \dfrac{.8}{.1}}{.191885} = 4.706 = -a_1;$$

$$h_2 = \frac{\log \dfrac{1-\beta}{\alpha}}{g_1+g_2} = \frac{\log \dfrac{.9}{.2}}{.191885} = 3.404 = a_2;$$

$$b = \frac{g_2}{g_1+g_2} = \frac{.066947}{.191885} = .3489;$$

$$D_1 = -4.706 + .3489N;$$

$$D_2 = 3.404 + .3489N.$$

Applying these equations we obtain the results shown in Table 2. Each value recorded for D_1 is the next integer smaller than the number obtained by the formula, while for D_2 it is the next integer larger. In this table decision numbers are shown only for selected values of N. Actually it would be possible to fail a student on the basis of six questions if he misses all of them, or to pass him on the basis of 14 questions all answered correctly. The grades shown are those of Table 1.

In case the examination ends before a statistical decision is reached within the framework of the sequential sampling plan, the student is given the grade indicate by Table 1.

TABLE 2

DECISION NUMBERS AND GRADES FOR SELECTED NUMBERS OF QUESTIONS

Number of Questions N	Number of errors permitted for passing D_1	Per cent of errors	Grade	Number of errors required for failing* D_2	Per cent of errors	Grade
20	2	10.0	A	11	55.0	F
40	9	22.5	B	18	45.0	F
60	16	26.7	B	25	41.7	F
80	23	28.8	C	32	40.0	F
100	30	30.0	C	39	39.0	F
200	65	32.5	D	73	36.5	E
1,000	344	34.4	D	353	35.3	E
∞	—	34.89	D	...	34.89	D

* If an *extremely* large number of questions were answered an "unfavorable" decision could conceivably be reached which would barely pass a student, since the proportion of errors approaches $b(=.3489)$ as N approaches ∞. By selecting a slightly higher value for p_0 and for p_1 a value of exactly .35 could be obtained for b, and an "unfavorable" decision would always fail a student. For instance, if $p_0=.3022$ and $p_1=.4$, or $p_0=.3$ and $p_1=.4073$, $b=.3500$.

TABLE 3

RESULTS OF A TRUE-FALSE EXAMINATION USING THE SEQUENTIAL METHOD OF SAMPLING

(After each name is indicated the number of errors, and the grade if a decision is reached)

N	20	40	60	80	Final Grade
D_1	2	9	16	23	
Rank					
1	Carr: 0	Carr: 5 (A)	—	...	Carr: A
2	Horn: 5	Holt: 11	Holt: 16 (B)	.	Holt: B
3	Abel: 6	Horn: 11	Abel: 17	.	Abel: C
4	Gill: 7	Swan: 11	Horn: 18	.	Horn: C
5	Hart: 7	Abel: 12	Swan: 18	.	Swan: C
6	Holt: 7	Gill: 13	Ford: 19	. .	Ford: C
7	Swan: 7	Hart: 14	Gill: 20	. .	Gill: C
8	Lamb: 8	Ford: 16	Lamb: 23	Lamb: 28	Lamb: D
9	Ford: 9	Lamb: 16	Hart: 25 (F)	. .	Hart: F
10	Sage: 11 (F)	—	—	.	Sage: F
D_2	11	18	25	32	. .

In the present case students were required to answer 20 questions each "round." The sets of 20 questions were not strictly random samples. For round one, all questions with numbers ending in 5 were

answered; for round 2 all questions with numbers ending in 0; and so on. It seems plausible that the effect of this method of sampling is to reduce the sampling variability, and therefore to make the probability of failing a good student somewhat less than α, and the probability of passing a poor student somewhat less than β.

The results of the experiment are shown (using fictitious names) in Table 3. Both of the students who went out on the unfavorable side decided to continue with the examination. One of them continued through 100 questions, but ended up with more than D_2 errors. The other voluntarily terminated the examination with 80 questions and managed to get barely into the indeterminate zone, but not far enough to avoid a grade of F. (Table 3 does not show voluntary continuations after a decision has been reached by sequential analysis.)

Two questions remain to be answered. (1) How many questions will be required, on the average, to reach a decision? (2) How close to ideal are the results for this particular sampling scheme?

The number of questions that tend to be required depends on the caliber of the student. Mediocre students will tend to require more questions than very good or very poor students. The expected number of question $\overline{N}$ for each of five different types of student can be ascertained easily by use of approximate formulas. This is done below.

Proportion of questions that will be missed in populations of questions p	*Average number of questions required* $\overline{N}$
0.0	$\overline{N}_0=\dfrac{h_1}{b}=\dfrac{4.706}{.3489}=13.5$
$p_1=.3$	$\overline{N}_{p_1}=\dfrac{(1-\alpha)h_1-\alpha h_2}{b-p_1}=\dfrac{.8(4.706)-.2(3.404)}{.3489-.3}=63.1$
$b=.3489$	$\overline{N}_b=\overline{N}_0\overline{N}_1=(13.488)(5.228)=70.5$
$p_2=.4$	$\overline{N}_{p_2}=\dfrac{(1-\beta)h_2-\beta h_1}{p_2-b}=\dfrac{.9(3.404)-.1(4.706)}{.4-.3489}=50.7$
1.0	$\overline{N}_1=\dfrac{h_2}{1-b}=\dfrac{3.404}{.6511}=5.2$

The next integer larger than $\overline{N}_0$ is the smallest sample size possible for passing a student, while the next integer larger than $\overline{N}_1$ is the smallest possible sample size for failing him. Since most of the students will tend to be C-grade, and will miss about 30 per cent of the questions,

about 63 questions should be required on the average. In the present instance it appears from Table 3 that the average number of questions required for a decision is somewhat larger than indicated by the formulas. It seems to the writer that this is at least partly accounted for by the non-random character of the sample, which should have the effect of reducing the sampling variability.[1] The average number of questions required reaches a maximum at a value for p that is close to b, $\bar{N}_b = 71$ in this case. In general it is advisable that the teacher be prepared to administer at least $2\bar{N}_b$ questions. If attainment of the desired objectives is possible only by use of more questions than it is practical to administer, one must revise his objectives. The number of questions required can be reduced by adopting values of p_1, p_2, α, and β so as to bring the values of a_1 and a_2 closer together. Methods of accomplishing this result have already been given.

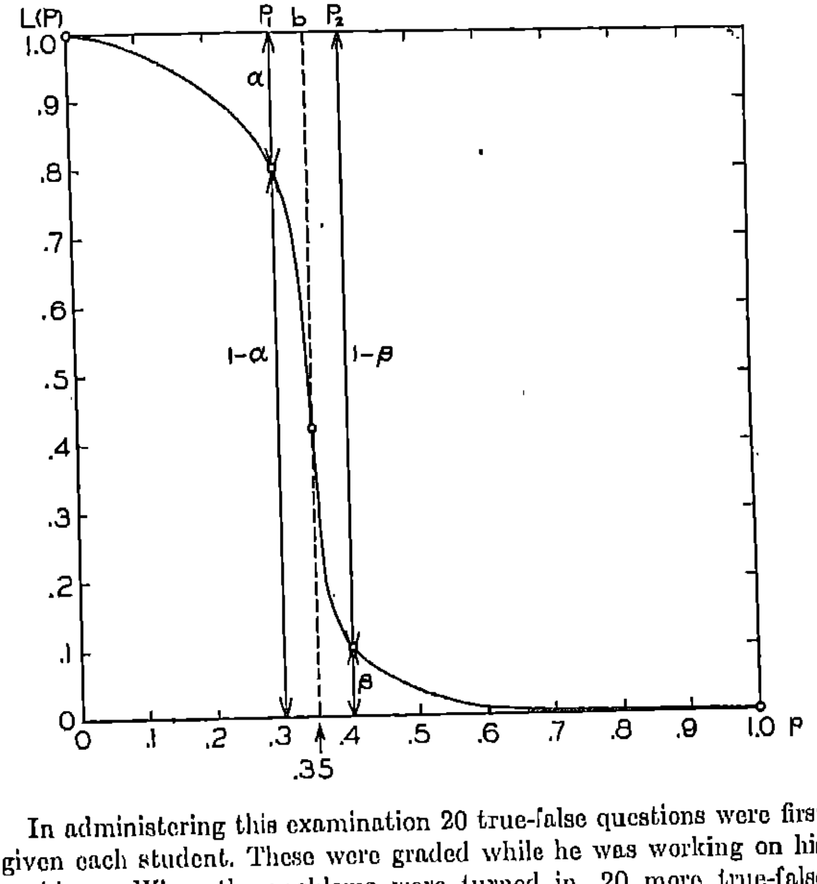
Comparison of the results of a given plan with ideal can be accomplished by plotting an *operating characteristic curve*. Such a curve is shown in the accompanying chart. The horizontal axis shows p the proportion of questions which the student tends to miss in the long run. This is roughly one-half the questions to which he doesn't know the answers, since he should guess about one-half of those correctly anyway. The vertical axis shows $L(p)$ the probability of passing the examination. Four points for this curve are already known, while a fifth can be computed very easily.

p	$L(p)$
0	1.00
p_1	$1-\alpha = .80$
b	$\dfrac{h_2}{h_1+h_2} = \dfrac{3.404}{4.700+3.404} = .42$
p_2	$\beta = .10$
1	0

[1]The true decision number lines can be brought closer together by increasing α and/or β, or by employing the principle of stratified sampling and thus reducing the sampling variability. If the sampling variability is reduced, but the D_1 and D_2 lines are not accordingly brought closer together too large a proportion of the points will tend to fall in the indeterminate zone between D_1 and D_2 until a fairly large sample is accumulated. To take an extreme illustration, where a student always runs true to form and there is therefore no sampling variability: (1) a decision will never be reached for a "borderline" student, while according to random sampling theory the average sample size for reaching a decision is 70.5; (2) a "good" student cannot fail, and can pass only by answering 100 questions and missing $.3 \times 100 = 30$ questions, as compared with an average sample size of 63.1 for random sampling; (3) a poor student cannot pass, and can fail only by answering 80 questions and missing $.4 \times 80 = 32$ questions, as compared with an average sample size of 60.7 for random sampling.

The chart shows a vertical broken line at $p = .35$ (which is almost the exact value of b). Ideally all students should be passed for whom $p < .35$, and all students should be failed for whom $p > .35$. The closer p_1 and p_2 are together, and the smaller the values of α and β, the closer the curve approaches to ideal; but also the larger the average number of questions that will be required.

OPERATING CHARACTERISTIC CURVE
($p_1 = .3$; $p_2 = .4$; $\alpha = .2$; $\beta = .1$)



In administering this examination 20 true-false questions were first given each student. These were graded while he was working on his problems. When the problems were turned in, 20 more true-false question were given, then a third set of 20 if necessary, and so on. Administrative difficulties encountered indicate that it would be better to give (say) 40 questions on the first round, with a diminishing number on subsequent rounds. It also seems that it would be better to

allow enough time for the examination to permit as many as 150 questions to be answered by marginal students. Another point to consider is whether it is better to give a random sample of questions each round, or a systematic stratified sample such as that used in this experiment. Since the variability from sample to sample is presumably smaller for the more representative systematic stratified sample, and therefore, apparently, the probability is reduced of unjustifiably passing or failing a student, it is the writer's opinion that a sample of the type selected is better. On the other hand, it might be worthwhile to work out some modification of the present formulas for use with procedures of the type we have described.

THE STATISTICAL SIGN TEST*

W. J. Dixon
University of Oregon
A. M. Mood
Iowa State College

This paper presents and illustrates a simple statistical test for judging whether one of two materials or treatments is better than the other. The data to which the test is applied consist of paired observations on the two materials or treatments. The test is based on the signs of the differences between the pairs of observations.

It is immaterial whether all the pairs of observations are comparable or not. However, when *all* the pairs are comparable, there are more efficient tests (the t test, for example) which take account of the magnitudes as well the signs of the differences. Even in this case, the simplicity of the sign test makes it a useful tool for a quick preliminary appraisal of the data.

In this paper the results of previously published work on the sign test have been included, together with a table of significance levels and illustrative examples.

INTRODUCTION

In experimental investigations, it is often desired to compare two materials or treatments under various sets of conditions. Pairs of observations (one observation for each of the two materials or treatments) are obtained for each of the separate sets of conditions. For example, in comparing the yield of two hybrid lines of corn, A and B, one might have a few results from each of several experiments carried out under widely varying conditions. The experiments may have been performed on different soil types, with different fertilizers, and in different years with consequent variations in seasonal effects such as rainfall, temperature, amount of sunshine, and so forth. It is supposed that both lines appeared equally often in each block of each experiment so that the observed yields occur in pairs (one yield for each line) produced under quite similar conditions.

The above example illustrates the circumstances under which the sign test is most useful:

(a) There are pairs of observations on two things being compared.

* This paper is an adaptation of a memorandum submitted to the Applied Mathematics Panel by the Statistical Research Group, Princeton University. The Statistical Research Group operated under a contract with the Office of Scientific Research and Development, and was directed by the Applied Mathematics Panel of the National Defense Research Committee.

(b) Each of the two observations of a given pair arose under similar conditions.

(c) The different pairs were observed under different conditions.

This last condition generally makes the t test invalid. If this were not the case (that is, if all the pairs of observations were comparable), the t test would ordinarily be employed unless there were other reasons, for example, obvious non-normality, for not using it.

Even when the t test is the appropriate technique many statisticians like to use the sign test because of its extreme simplicity. One merely counts the number of positive and negative differences and refers to a table of significance values. Frequently the question of significance may be settled at once by the sign test without any need for calculations.

It should be pointed out that, strictly speaking, the methods of this paper are applicable only to the case in which no ties in paired comparisons occur. In practice, however, even when ties would not occur if measurements were sufficiently precise, ties do occur because measurements are often made only to the nearest unit or tenth of a unit for example. Such ties should be included among the observations with half of them being counted as positive and half negative.

Finally, it is assumed that the differences between paired observations are independent, that is, that the outcome of one pair of observations is in no way influenced by the outcome of any other pair.

PROCEDURE

Let A and B represent two materials or treatments to be compared. Let x and y represent measurements made on A and B. Let the number of pairs of observations be n. The n pairs of observations and their differences may be denoted by:

$$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$$

and

$$x_1 - y_1, x_2 - y_2, \cdots, x_n - y_n.$$

The sign test is based on the signs of these differences. The letter r will be used to denote the number of times the less frequent sign occurs. If some of the differences are zero, half of them will be given a plus sign and half a minus sign.

As an example of the type of data for which the sign test is appropriate, we may consider the following yields of two hybrid lines of corn obtained from several different experiments. In this example $n = 28$ and $r = 7$.

If there is no difference in the yielding ability of the two lines, the positive and negative signs should be distributed by the binomial distribution with $p=\frac{1}{2}$. The null hypothesis here is that each difference has a probability distribution (which need not be the same for all differences) with median equal to zero. This null hypothesis will obtain, for instance, if each difference is symmetrically distributed about a mean of zero, although such symmetry is not necessary. The null hypothesis will be rejected when the numbers of positive and negative signs differ significantly from equality.

YIELDS OF TWO HYBRID LINES OF CORN

Experiment Number	Yield of A	Yield of B	Sign of $x-y$	Experiment Number	Yield of A	Yield of B	Sign of $x-y$
1	47.8	46.1	+	4	40.6	41.3	−
	48.6	50.1	−		39.8	40.8	−
	47.6	48.2	−		42.2	42.0	+
	43.0	48.6	−		41.4	42.5	−
	42.1	43.4	−				
	41.0	42.9	−	5	38.9	39.1	−
					39.0	39.4	−
2	28.0	33.6	−		37.5	37.3	+
	29.0	31.1	−				
	27.4	28.0	−	6	36.8	37.5	−
	28.1	27.5	+		35.9	37.3	−
	28.0	28.7	−		33.6	34.0	−
	28.3	28.8	−				
	26.4	26.3	+	7	39.2	40.1	−
	26.8	26.1	+		39.1	42.6	−
3	33.3	32.4	+				
	30.6	31.7	−				

Table 1 gives the critical values of r for the 1, 5, 10, and 25 per cent levels of significance. A discussion of how these values are computed may be found in the appendix. A value of r less than or equal to that in the table is significant at the given per cent level.

Thus in the example above where $n=28$ and $r=7$, there is significance at the 5% level, as shown by Table 1. That is, the chances are only 1 in 20 of obtaining a value of r equal to or less than 8 when there is no real difference in the yields of the two lines of corn. It is concluded, therefore, at the *5% level of significance*, that the two lines have different yields.

In general, there are no values of r which correspond exactly to the levels of significance 1, 5, 10, 25 per cent. The values given are such that they result in a level of significance as close as possible to, but not exceeding 1, 5, 10, 25 per cent. Thus, the test is a little more strict,

TABLE 1

TABLE OF CRITICAL VALUES OF r FOR THE SIGN TEST

n	Per Cent Level of Significance 1	5	10	25	n	Per Cent Level of Significance 1	5	10	25
1	—	—	—	—	51	15	18	19	20
2	—	—	—	—	52	16	18	19	21
3	—	—	—	0	53	16	18	20	21
4	—	—	—	0	54	17	19	20	22
5	—	—	0	0	55	17	19	20	22
6	—	0	0	1	56	17	20	21	23
7	—	0	0	1	57	18	20	21	23
8	0	0	1	1	58	18	21	22	24
9	0	1	1	2	59	19	21	22	24
10	0	1	1	2	60	19	21	23	25
11	0	1	2	3	61	20	22	23	25
12	1	2	2	3	62	20	22	24	25
13	1	2	3	3	63	20	23	24	26
14	1	2	3	4	64	21	23	24	26
15	2	3	3	4	65	21	24	25	27
16	2	3	4	5	66	22	24	25	27
17	2	4	4	5	67	22	25	26	28
18	3	4	5	6	68	22	25	26	28
19	3	4	5	6	69	23	25	27	29
20	3	5	5	6	70	23	26	27	29
21	4	5	6	7	71	24	26	28	30
22	4	5	6	7	72	24	27	28	30
23	4	6	7	8	73	25	27	28	31
24	5	6	7	8	74	25	28	29	31
25	5	7	7	9	75	25	28	29	32
26	6	7	8	9	76	26	28	30	32
27	6	7	8	10	77	26	29	30	32
28	6	8	9	10	78	27	29	31	33
29	7	8	9	10	79	27	30	31	33
30	7	9	10	11	80	28	30	32	34
31	7	9	10	11	81	28	31	32	34
32	8	9	10	12	82	28	31	33	35
33	8	10	11	12	83	29	32	33	35
34	9	10	11	13	84	29	32	33	36
35	9	11	12	13	85	30	32	34	36
36	9	11	12	14	86	30	33	34	37
37	10	12	13	14	87	31	33	35	37
38	10	12	13	14	88	31	34	35	38
39	11	12	13	15	89	31	34	36	38
40	11	13	14	15	90	32	35	36	39
41	11	13	14	16	91	32	35	37	39
42	12	14	15	16	92	33	36	37	39
43	12	14	15	17	93	33	36	38	40
44	13	15	16	17	94	34	37	38	40
45	13	15	16	18	95	34	37	38	41
46	13	15	16	18	96	34	37	39	41
47	14	16	17	19	97	35	38	39	42
48	14	16	17	19	98	35	38	40	42
49	15	17	18	19	99	36	39	40	43
50	15	17	18	20	100	36	39	41	43

For $n > 100$, approximate values of r may be found by taking the nearest integer less than $\frac{1}{2}(n - k\sqrt{n})$, where $k = 1.3, 1, .82, .58$ for the 1, 5, 10, 25 per cent values respectively. A closer approximation to the values of r is obtained from $\frac{1}{2}(n-1) - k\sqrt{n+1}$ and the *more exact values of* k, 1.2879, .9800, .8224, .5752.

on the average, than the level of significance which is indicated. For small samples the test is considerably more strict in some cases. For example, the value of r for $n=12$ for the 10 per cent level of significance actually corresponds to a per cent level less than 5.

The critical values of r in Table 1 for the various levels of significance were computed for the cases where either the $+$'s or $-$'s occur a significantly small number of times. Sometimes the interest may be in only one of the signs. For example, in testing two treatments, A and B, A may be identical with B except for certain additions which can only have the effect of improving B. In this case one would be interested only in whether the deficiency of minus signs (for differences in the direction A minus B) were significant or not. In cases of this kind the per cent levels of significance in Table 1 would be divided by two. Thus, 8 minus signs in a sample of 28 would correspond to the 2.5% level of significance.

SIZE OF SAMPLE

Even though there is no real difference, a sample of four or even five with all signs alike will occur by chance more than 5% of the time. Four signs alike will occur by chance 12.5% of the time and five signs alike will occur by chance 6.25% of the time. Therefore, at the 5% level of significance, it is necessary to have at least six pairs of observations even if all signs are alike before any decision can be made. As in most statistical work, more reliable results are obtained from a larger number of observations. One would not ordinarily use the sign test for samples as small as 10 or 15, except for rough or preliminary work.

The question may be raised as to the minimum sample size necessary to detect a given difference in two materials. Suppose that in an indefinitely large number of observations 30% $+$'s and 70% $-$'s are to be expected and that we wish the sample to be large enough to detect this difference at the 1% level of significance. Although no sample, however large, will make it absolutely certain that a significant difference will be found, the sample size can be chosen to make the probability of finding a significant result as near to certainty as is desired. In Table 2, this probability has been chosen as 95%; the minimum values of n (sample size) and the corresponding critical values of r to insure a decision 95% of the time are given for various actual percentages p_0 and levels of significance α.

The sign test merely measures the significance of departures from a 50–50 distribution. If the signs are actually distributed 45–55, then the departure from 50–50 is not likely to be significant unless the

sample is quite large. Table 2 shows that if the signs are actually distributed 45–55, then one must take samples of 1,297 pairs in order to get a significant departure from a 50–50 distribution at the 5% level of significance. The number 1,297 is selected to give the desired significance 95% of the time; that is, if a large number of samples of 1,297 each were drawn from a 45–55 distribution, then 95% of those samples could be expected to indicate a significant departure (at the 5% level) from a 50–50 distribution.

TABLE 2

MINIMUM VALUES OF n NECESSARY TO FIND SIGNIFICANT DIFFERENCES 95% OF THE TIME FOR VARIOUS GIVEN PROPORTIONS

p_1	n				r			
	α =1%	5%	10%	25%	α =1%	5%	10%	25%
.45 (.55)	*1,777*	*1,297*	*1,080*	*780*	*813*	*618*	*513*	*375*
.40 (.60)	*443*	327	*267*	*193*	*193*	145	*110*	*87*
.35 (.65)	*193*	143	*118*	86	*78*	59	*49*	37
.30 (.70)	*100*	79	67	47	*39*	30	24	19
.25 (.75)	66	49	42	32	22	17	15	12
.20 (.80)	44	35	28	21	13	11	9	7
.15 (.85)	32	23	18	14	8	6	5	4
.10 (.90)	24	17	13	11	6	4	3	3
.05 (.95)	15	12	11	6	2	2	2	1

The italicised values are approximate. The maximum error is about 5 for the value of n, and 2 for the value of r. The values of n and r for 5% were taken from MacStewart (reference 1) who gives a table of values of n and r for a range of confidence coefficients (the above table uses only 95%) and a single value α =5%.

Of course, in practice one would not do any testing if he knew in advance the expected distribution of signs (that it was 45–55, for example). The practical significance of Table 2 is of the following nature: In comparing two materials one is interested in determining whether they are of about equal or of different value. Before the investigation is begun, a decision must be made as to how different the materials must be in order to be classed as different. Expressed in another way, how large a difference may be tolerated in the statement that "the two materials are of about equal value?" This decision, together with Table 2, determines the sample size. If one is interested in detecting a difference so small that the signs may be distributed 45–55, he must be prepared to take a very large sample. If, however, one is interested only in detecting larger differences, (for example, differences represented by a 70–30 distribution of signs), a smaller sample will suffice.

In many investigations, the sample size can be left undetermined,

and only as much data accumulated as is needed to arrive at a decision. In such cases, the sign test could be used in conjunction with methods of sequential analysis. These methods provide a desired amount of information with the minimum amount of sampling on the average. A complete exposition of the theory and practice of sequential analysis may be found in references 3 and 4.

MODIFICATIONS OF THE SIGN TEST

When the data are homogeneous (measurements are comparable between pairs of observations), the sign test can be used to answer questions of the following kind:

1. Is material A better than B by P per cent?
2. Is material A better than B by Q units?

The first question would be tested by increasing the measurement on B by P per cent and comparing the results with the measurements on A. Thus, let

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \text{ etc.}$$

be pairs of measurements on A and B, and suppose one wished to test the hypothesis that the measurements, x, on A were 5% higher than the measurements, y, on B. The sign test would simply be applied to the signs of the differences

$$x_1 - 1.05y_1, x_2 - 1.05y_2, x_3 - 1.05y_3, \text{ etc.}$$

In the case of the second question the sign test would be applied to the differences

$$x_1 - (y_1 + Q), x_2 - (y_2 + Q), x_3 - (y_3 + Q), \text{ etc.}$$

In either case, if the resulting distribution of signs is not significantly different from 50-50, the data are not inconsistent with a positive answer to the question. Usually there will be a range of values of P (or Q) which will produce a non-significant distribution of signs. If one determines such a range, using the 5% level of significance for example, then that range will be a 95% confidence interval for P (or Q).

Even when the data are not homogeneous, it may be possible to frame questions of the above kind, or it may be possible to change the scales of measurement so that such questions would be meaningful.

MATHEMATICAL APPENDIX

A. *Assumptions*. Let observations on two materials or treatments A and B be denoted by x and y, respectively. It is assumed that for any pair of observations (x_i, y_i) there is a probability $p(0 < p < 1)$ that

$x_i > y_i$ ($i=1, 2, \cdots, n$); p is assumed to be unknown.[1] It is also assumed that the n pairs of observations (x_i, y_i), ($i=1, 2, \cdots, n$) are independent; i.e., the outcome (+ or −) for (x_j, y_j) is independent of the outcome for (x_i, y_i) ($i \neq j$).

B. *The Observations.* The purpose of obtaining observations (x_i, y_i) is to make an inference regarding p. The observed quantity upon which an inference is to be based is r, the number of +'s or −'s (whichever occur in fewer numbers) obtained from n paired observations (x_i, y_i). On the basis of the assumption above it follows that the probability of obtaining exactly r as the minimum number of +'s or −'s is:

$$\binom{n}{r}\left[p^r(1-p)^{n-r} + p^{n-r}(1-p)^r\right] \quad r = 0, 1, 2, \cdots, \frac{n-1}{2};\ n \text{ odd}$$

$$r = 0, 1, 2, \cdots, \frac{n-2}{2};\ n \text{ even}$$

$$\binom{n}{\frac{1}{2}n} p^{\frac{1}{2}n}(1-p)^{\frac{1}{2}n} \qquad r = \frac{n}{2};\ n \text{ even}.$$

C. *The Inference.* In the sign test the hypothesis being tested is that $p=\frac{1}{2}$; in other words that the distributions of the differences $x_i - y_i$ ($i=1, 2, \cdots, n$) have zero medians. For the more general tests discussed in Section 5, the hypothesis is that the differences $x_i - f(y_i)$ ($i=1, 2, \cdots, n$) have zero medians. The function $f(y)$ may be Py or $Q+y$ (where P and Q are the constants mentioned in Section 5) or any other function appropriate for comparison with x in the problem at hand.

The hypothesis that $p=\frac{1}{2}$ is tested by dividing the possible values of r into two classes and accepting or rejecting the hypothesis according as r falls in one or the other class. The classes are chosen so as to make small (say $\leq \alpha$) the chance of rejecting the hypothesis when it is true and also to make small the chance of accepting the hypothesis when it is untrue. It can be shown that in a certain sense, the best set of rejection values for r is $0, 1, \cdots, R$, where R depends on α and n. R can be determined by solving for $R=$ maximum i in the inequality:

$$\sum_{j=0}^{i}\binom{n}{j}\left(\frac{1}{2}\right)^n = I_{\frac{1}{2}}(n-i, i+1) \leq \tfrac{1}{2}\alpha$$

where $I_x(a, b)$ is the incomplete beta function. Table 1 was computed in this way.

[1] An additional assumption is that the probability $A_i = B_i$ is 0; thus the probability $B_i > A_i$ is $(1-p)$.

D. *Sample Sizes.* When the sample size is small the sign test is likely to reject the hypothesis, $p=\frac{1}{2}$, only if p is near zero or one. If p is near, but not equal to $\frac{1}{2}$, the test is likely to reject the hypothesis, $p=\frac{1}{2}$, only when the sample is large.

The sample size required to reject the hypothesis $p=\frac{1}{2}$ at the α level of significance, $100\lambda\%$ of the time, may be determined by finding the largest i and smallest n which satisfy:

$$\sum_{j=0}^{i} \binom{n}{j} \left(\frac{1}{2}\right)^n \leqq \frac{1}{2}\alpha$$

and

$$\sum_{j=0}^{i} \binom{n}{j} p^j(1-p)^{n-j} \geqq \lambda \qquad\qquad p < \tfrac{1}{2}.$$

n and i are given in Table II for various values of p and α; λ was taken to be .95 in all cases. The tabular values for $1-p$ are the same as those for p because of the symmetry of the binomial distribution.

E. *Efficiency of the Sign Test.* Let $z=x-y$. Assume z is normally distributed with mean a and variance σ^2. The probabality of obtaining a $+$ on a particular z_i is:

$$p = \frac{1}{\sqrt{2\pi}} \int_{-a/\sigma}^{\infty} e^{-\frac{1}{2}u^2} du.$$

An estimate of p involving only the signs of z_i $(i=1, 2, \cdots, n)$ yields an estimate of (a/σ). Cochran (reference 2) has shown that in large samples the variance of this estimate of (a/σ) is $2\pi pq\, e^{(a/\sigma)^2}/n$. We shall denote a/σ by c.

The efficiency of an estimate based on n independent observations is defined as the limit (as $n\to\infty$) of the ratio of the variance of an efficient estimate to that of the given estimate. An efficient estimate of c is:

$$\frac{\bar{z}}{\sqrt{\dfrac{\sum (z_i - \bar{z})^2}{n-1}}} = \frac{t}{\sqrt{n}}$$

where t is Student's t and $\bar{z} = \sum z_i/n$.

The variance of this estimate is $1/(n-2)$; thus the efficiency, E, of the sign test is $e^{-c^2}/2\pi pq$. If $c=0$, then $p=\frac{1}{2}$ and the efficiency is $2/\pi = 63.7\%$.

The preceding discussion pertains to large values of n; for small values of n, the efficiency is a little better than 63.7%. Computations

were made for several smaller values of n, namely, for $n=18$, 30, 44 pairs of observations at the 10% level of significance. It was found that the sign test using 18 pairs of observations is approximately equivalent to the t-test using 12 pairs of observations; for 30 pairs the equivalent t-test requires between 20 and 21 pairs; and for 44 pairs the equivalent t-test requires between 28 and 29 pairs. Cochran shows that the efficiency of r/n for estimating c decreases as $|c|$ increases.

REFERENCES

[1] W. MacStewart, "A note on the power of the sign test," *Annals of Mathematical Statistics*, Vol. 12 (1941), pp. 236–238.

[2] W. G. Cochran, "The efficiencies of the binomial series test of significance of a mean and of a correlation coefficient," *Journal Royal Statistical Society*, Vol. C, Part I (1937), pp. 69–73.

[3] A. Wald, "Sequential method of sampling for deciding between two courses of action," *Journal American Statistical Association*, Vol. 40 (1945), pp. 277–306.

[4] Statistical Research Group, Columbia University, *Sequential Analysis of Statistical Data: Applications* (1945), Columbia University Press, New York.

TESTS OF INCREASED SEVERITY

C. West Churchman
University of Pennsylvania
Benjamin Epstein
Coal Research Laboratory, Carnegie Institute of Technology

Certain manufactured items must be produced of such quality that only a very small fraction will fail when used. In problems of this sort standard sampling plans are impractical because of the excessively large sample sizes required. In this paper we consider the statistical questions involved in developing an efficient sampling plan for such problems, and recommend a procedure which has been found very useful in biological assay and in the testing of explosives. There is good reason to believe that the same statistical techniques can also be used in many other fields in which one is interested in studying the response of objects tested as a function of the strength of the stimuli to which they are subjected.

I. INTRODUCTION

There are numerous manufactured items which must be so constructed that only a very small fraction of these items will fail to perform under certain operating conditions. For example: (a) a munitions plant is required to produce primers sufficiently sensitive to ensure that very few and preferably no primers will fail to explode under a specified blow; (b) an aircraft manufacturer may want to be certain that the strength of welds meets some specified requirement (e.g., one might require that only 1 in 1,000 welds should fail to meet shear stress of X pounds); (c) a rope manufacturer may wish to make a product which will be able to withstand a steady load of X pounds applied for Y minutes; (d) the manufacturer of electrical equipment may desire to make a product capable of withstanding some specified voltage with virtually no breakdown of the electrical insulation; (e) a drug manufacturer may wish to test the effectiveness of a drug in destroying a certain organism.

The data arising from all problems of this type may be termed "Sensitivity Data." This is a general term applied to that type of experimental data for which the act of taking a measurement on the object tested either destroys the object or so changes the properties of the object that it can no longer be used in further testing.

We shall call the "stimulus" that condition to which we subject the samples, and we shall call the stimulus at which a given object just

fails the "critical stimulus"; that is, for a stimulus of less intensity than the critical stimulus, the object will not fail, and for a stimulus of greater intensity than the critical stimulus, the object will fail. These terms are illustrated in the following examples:

Objects Tested	*Type of Stimulus*	*Critical Stimulus*
Primer Caps	Blow of a certain energy	Energy of blow just necessary to explode
Welds	Shear stress (in pounds)	Stress just necessary to shear
Rope	Load (in pounds)	Load just necessary to break
Insulation	Voltage	Voltage just necessary to give breakdown
Organisms	Dosage of drug	Amount just necessary to kill

It is the purpose of this paper to give the general mathematical and statistical attack recently developed for problems of this nature and to compare this approach with other current methods. Sensitivity problems may be divided into two broad categories:

I. The applied blow or stimulus or particular test condition will result either in failure or non-failure of the specimen tested. A numerical value cannot be attached to the result of any test. The specimen either fails or passes, explodes or does not explode, breaks or does not break.

II. The stimulus is so applied that each object tested has associated with it a precise stimulus number (i.e., a value of the stimulus for which the object responds critically). For example, to any weld X_i there can be associated a shear stress Y_i.

In this paper we shall restrict ourselves to problems belonging to category I because the data associated with problems in category II may be readily analyzed by the use of known techniques in the elementary theory of frequency distributions and curve fitting.[1]

It will be our basic aim in this paper to show how one can determine the distribution of stimuli for which the objects will respond critically, even though it is impossible to measure the exact value of the stimulus at which a given specimen will respond critically. For example, the physiologist is interested in the value of the dosage for which 50% of the specimens tested will survive. Yet, in practice, one cannot arrive at the precise value of the critical stimulus for each specimen by a method of successive approximations, since once an animal has been injected with any concentration of the drug at all, it can no longer be used for further tests.[2]

[1] As an example of a problem belonging to category II see an article by C. R. Barrett, "Quality Control with Sampling Inspection," *Mechanical Engineering*, 64, 301-304 (1942).

[2] See Statistical Note No. 1.

The particular use of sensitivity data which we will consider for the moment will be the problem whether or not a lot meets the requirement that only a small fraction (.001 or .0001 for example) fails under certain operating conditions; the inference is to be made from a sample. A number of methods can be used in testing such a hypothesis. Preference for one method over another will be based on the principle that we desire the *maximum* of reliable information from the minimum of data, plus the amount of prior information we can bring to bear.

At present there are four methods available for analyzing sensitivity data belonging to category I. These are:

I(a). Tests conducted only at the stimulus required by the specifications. The submitted lot may be judged on the basis of the behavior of a sample at the specification stimulus;

I(b). Tests conducted at some stimulus which subjects the sample to more stringent operating conditions (e.g., to heavier blows, greater stresses, smaller or larger doses of a drug, etc.);

I(c). Tests conducted at two and sometimes three stimuli, all of them being generally different from the specification stimulus;

I(d). Complete run-down tests, i.e., tests conducted over a discrete series of stimuli from the stimulus resulting in no failures to the stimulus yielding 100% failures.

II. TESTS BELONGING TO CATEGORY I(a)

Tests of this type are usually inefficient, since one cannot get good assurance that a submitted lot will have the high quality required by the specification, even if all members in the small sample pass the test. In order to get such assurance huge samples are required.

To bring these points into bold relief, let us consider the following: what can be said about the quality of a lot, if a random sample of 20 or even 100 specimens taken from the lot contains no defectives?

The answer to this question is given by the following table[3] which gives the probability of accepting material of a given quality on the basis of samples of 20 or 100 if the inspection plan consists of accepting if no defects occur, and rejecting otherwise. (Of course, if retests are permitted the probability of acceptance will increase.)

For example, the end values in the second line of Table 1 may be read as follows: if acceptance is based solely on no defects in 20, then the chances are 1 in 100 of accepting material 21% defective and 99 in 100 of accepting material .05% defective. Thus, a sample of 20, or

[3] See Statistical Note No. 2.

even 100, is virtually useless if one wishes to accept very few lots of a quality level worse than .1% defective, say.

The following general remarks may be made about this kind of test: (1) its *advantage* is that it makes no presupposition about the lot and

TABLE 1

QUALITY OF LOTS THAT MAY BE ACCEPTED CORRESPONDING TO VARIOUS PROBABILITIES IF NO DEFECTIVES ARE ALLOWED IN THE SAMPLE

Sample Size	Probability of Acceptance						
	.01	.05	.1	.5	.9	.95	.99
20	21 %	14 %	11 %	3.4%	.5%	.25%	.05%
100	4.5%	2.9%	2.3%	.7%	.1%	.05%	.01%

the method of inspection except random (representative) sampling; (2) its *disadvantages* are that for even a moderate degree of assurance, large samples are needed, and further, without additional information, one cannot extrapolate from the inspection test to all the many conditions to be expected in service.

III. TESTS BELONGING TO CATEGORY I(b)

The disadvantages of tests of type I(a) lead directly to tests falling into category I(b). Why should one carry out the tests at that stimulus level which is to yield less than .1% defective? Why not carry out the test under test conditions which are so severe that one may expect a larger percentage of the items (20% or 50%, for example) to fail the test? This idea is certainly not new. Any accelerated laboratory test accomplishes a similar purpose; test conditions are so altered that a lot may be graded more rapidly and with less work.

In general, the stimulus which should give approximately 50% failures in the lot is determined from a series of initial tests conducted on samples drawn from a lot which is chosen as standard.[4] A number of fairly large samples (in case of primer lots, 300 to 500; in case of drugs, perhaps 100 animals) are taken at random from the lot. Each of these large samples is broken up into a number of smaller sub-samples, each sub-sample being tested at a different stimulus level. From the observed set of p_i, the fraction failing to survive at the stimulus x_i, it is possible to obtain an estimate of the distribution of critical stimuli. In particular that stimulus for which one-half of the total lot fails to

[4] There is a certain degree of arbitrariness attached to the selection of a standard lot. Presumably the lot designated as standard is chosen because it gives satisfactory performance and is known to meet the requirements of the specification.

survive can be estimated. For convenience this is the stimulus used in further tests. The principle involved in obtaining a calibration distribution is an important one. It is the forerunner of the modern method of attack on sensitivity problems. When this approach is fully developed, it leads directly to the completely rundown in category I(d).

One should note carefully that a fundamental distribution assumption is made when lot quality is judged purely on the basis of the results obtained at the stimulus giving 50% failures in the reference lot. Increased severity tests carried out at a single stimulus, giving 50% failures in the reference lot, are sensitive indicators of significant differences among lots at the specification stimulus, only if one makes a certain assumption about the distribution of critical stimuli; for example, *that the "critical" stimuli are normally distributed with a known standard deviation which is the same for all lots.* If an assumption of this kind is justified, then a sensitive test for significance can be made even with small samples.

In order to get some idea of the efficiency of an increased severity test, consider the following example: (a) we shall call acceptable any lot which has .1% failures or fewer at a stimulus giving .1% failures in a standard lot; (b) we shall call unacceptable any lot which has 1% failures or more at the stimulus just mentioned; (c) we shall call marginal those lots which give between .1% and 1% failures; (d) the risk of rejecting a lot having .1% failures shall be .05; (e) the risk of accepting a lot having 1% failures shall be .01.[5] Then, it can be shown that even under the best possible sampling methods we shall be required on the average to take a sample of 628 specimens for lots giving .1% failures and a sample of 200 specimens for lots giving 1% failures in order to reach a decision.[6]

Now, suppose that we were to translate the problem involving the discrimination between lots which give .1% and 1% failures respectively at the stimulus giving .1% failures in the standard lot into one involving a comparison carried out at a stimulus giving 50% failures in the standard lot, then what sample sizes should be taken in order to arrive at a decision with the risks stated above?

One can easily construct an increased severity analogue to the problem just stated. The fundamental assumption that the initial stimuli are normally distributed with the same standard deviation means that the critical stimulus causing 50% failures in a lot having .1% failures at the specification stimulus will cause 78% failures in a lot having 1%

[5] In statistical terminology, the producer's risk $\alpha = .05$, and the consumer's risk $\beta = .01$.

[6] For details of such calculations see "Sequential Analysis of Statistical Data, Applications," *AMP Report* 30.2R, Section 2, revised.

failures at the specification stimulus?[7] Graphically the situation may be visualized as shown in Fig. 1, where Curve A is Distribution of Critical Stimuli in lots having .1% failures at the specification stimulus and Curve B is Distribution of Critical Stimuli in lots having 1% failures at the specification stimulus. Thus, a comparison between lots having .1% and 1% failures at the specification stimulus has been converted into one between lots having 50% and 78% failures, respectively, at the increased severity stimulus with producer's and consumer's risks equal to .05 and .01, respectively.

It can be shown that we shall be required to take on the average a sample of 22 specimens for lots having 50% failures at the increased severity level and a sample of 18 specimens for lots having 78% failures at this level in order to reach a decision. This means that reliable inferences may be drawn with greatly reduced samples if the tests are conducted at the increased severity level in the manner described.

The following general remarks may be made about tests in I(b): (1) the *advantage* is that it is possible to predict failures under service conditions with a relatively small sample; (2) the *disadvantages* are that very strong presuppositions must be made which can only be checked after considerable experimentation; in particular, not only must the form of the distribution function be known, but also the standard deviation. The latter is usually very difficult to estimate with any satisfactory degree of precision for this kind of test since even small errors in estimate will result in large errors in prediction.

IV. TESTS BELONGING TO CATEGORY I(c)

In practice it is found that not only does the value of the stimulus giving 50% failures vary from lot to lot, but so may the standard deviation of the critical stimuli. If this condition prevails, then an increased severity test conducted at only one point will give misleading and erroneous results. Such tests are inadequate for detecting differences among means due to the fact that the standard deviation may differ significantly from lot to lot. This will mean that the distribution curves associated with each lot will no longer be capable of transformation into one another by translation alone.

Recognition of the fact that the standard deviation of the critical stimuli plays a fundamental role in predicting failure under various

[7] *As a consequence of the fundamental distribution assumptions, reference to the table of areas* under the normal curve will show that if lot A has .1% failures at the specification stimulus and lot B has 1% failures at this stimulus, then the stimuli giving 50% failures in lots A and B will be 3.09σ and 2.33σ away, respectively, from the specification stimulus. This means that the stimulus giving 50% failures in lot A is separated by $.76\sigma$ from the stimulus giving 50% failures in lot B. Therefore, the stimulus giving 50% failures in lot A will give 78% failures in lot B.

conditions has led to tests belonging to category I(c), i.e., increased severity tests employing two and possibly three levels. In the case where the test is conducted at two stimuli x_1 and x_2, it is customary to choose x_1 as that stimulus for which about 20% of the objects tested may fail to pass and x_2 as that stimulus for which 80% of the objects tested may fail to pass.[8] π_i, the true fraction passing at stimulus x_i, may be represented as the integral of a normal distribution from x_i to ∞, i.e., as:

$$\pi_i = \int_{x_i}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx. \tag{1}$$

The problem is to determine estimates, $\bar{x}$ and s, of the true mean, μ, and true standard deviation, σ, from the experimentally determined values of π_1 and π_2 (denoted by p_1 and p_2) at the stimuli x_1 and x_2. Schematically, we have:

p_i = observed fraction passing at x_i
π_i = true fraction passing at x_i
$\bar{x}$ = sample estimate of the true mean critical stimulus
μ = true mean critical stimulus
s = sample estimate of the true standard deviation
σ = true standard deviation

This can readily be accomplished for a normal distribution by treating the experimentally determined p_1 and p_2 as the true π_1 and π_2 and associating with each p_i a t_i, where $t_i = x_i - \bar{x}/s$ and where values of t_i vs. p_i are to be found in any standard table of normal areas (e.g., see H. C. Carver, "Statistical Tables," Edwards Bros., 1941). In particular, we obtain the simultaneous linear equations:

$$t_1 = \frac{x_1 - \bar{x}}{s} \qquad t_2 = \frac{x_2 - \bar{x}}{s} \tag{2}$$

which may easily be solved for $\bar{x}$ and s. The solutions are:

$$\bar{x} = x_1 - \frac{dt_1}{t_2 - t_1} \quad \text{and} \quad s = +\frac{d}{t_2 - t_1}, \quad \text{where } d = x_2 - x_1. \tag{3}$$

[8] Actually, the best estimates are made if the percentages are 6% and 94%, respectively, in the sense that the sampling variance of the standard deviation is thereby minimized (see J. H. Gaddum, "Methods of Biology Assay," etc., *Medical Research Council Reports on Biological Standards*, Spec. Rep. Ser. No. 183); but for most procedures it is risky to try for these values since one is apt to obtain 0% or 100% values, and the method cannot then be applied. Consequently, whenever possible, one should aim at the compromise of 20% and 80%.

This method for determining estimates of the parameters of the normal distribution is a special case of the "probit" method. It also should be noted that as a direct consequence of carrying out the test at two levels we can compare two lots not only as to their means, but also as to their standard deviation or more generally as to $\mu \pm k\sigma$.

There may be advantages in using a third stimulus midway between the other two stimuli. Such a procedure permits at least a crude check on the assumption of normality by requiring that the middle stimulus produce a p_3 which in turn will give a (t_3, x_3) which is "close" to the line joining the points (t_1, x_1) and (t_2, x_2). To determine how closely the (t_i, x_i) should cluster about the line, one would have to run a conventional regression analysis with weighted coefficients.

In this connection, it should be noted that if a sensitivity test is conducted at more than two levels, then the simple equations (2) no longer suffice to determine $\hat{x}$ and s. In the more general case where one has more than two number pairs (p_i, x_i), one determines $\hat{x}$ and s from a least squares linear fit to the associated number pairs (t_i, x_i). This method is cumbersome, however, since the t_i do not have equal standard errors.[9]

The procedure for the two-stimuli test may be summarized as follows: (1) test a given size sample at two stimuli x_1 and x_2, preferably, though not necessarily, where the respective percentages are about 20 and 80; e.g., suppose $x_1 = 3$ foot-pounds, $x_2 = 5$ foot-pounds in a test where the stimulus is the energy of blow; (2) call the fraction unaffected at x_1, p_1, and the fraction unaffected at x_2, p_2; e.g., say p_1 is .84 and p_2 is .10; (3) in tables of areas of the normal curve (see any standard statistics text), find that value of t for which the area is p_1 (call it t_1), and that value of t for which the area is p_2; e.g., for $p_1 = .84$, $t_1 = -1.00$, and for $p_2 = .10$, $t_2 = +1.28$; (4) solve for $\hat{x}$ and s by the equations (3) given above; eg., in the example cited:

$$\hat{x} = 3 + \frac{2(1)}{2.28} = 3.88$$

$$s = \frac{2}{2.28} = .88;$$

[9] For a complete account of the application of the method to certain biological problems, see C. I. Bliss, "The Calculation of the Dosage Mortality Curve," *Annals of Applied Biology*, 22, 134-167. It is to be noted that in many cases where the critical stimuli are not normally distributed the assumption of normality may be satisfied, if the stimuli are subjected to certain familiar transformations, e.g., the logarithmic. Also, the complexities involved in Bliss's method may be avoided in part by employing distribution functions that are practically identical with the normal, or appear to fit the data better; e.g., the so called "growth curve"; see E. B. Wilson and H. Worcester, "The Determination of L.D. 50 and its Sampling Error in Bio-Assay," *Proceedings National Academy of Science*, 29, 1943.

(5) then, as a result of these calculations, one can predict that about nine failures in 400 will occur at $\bar{x}+2s$, about four failures in three thousand will occur at $\bar{x}+3s$, about one failure in 100,000 will occur at $\bar{x}+4s$, and about three failures in 10,000,000 will occur at $\bar{x}+5s$.[10]

The following general remarks may be made concerning tests in I(c): (1) the *advantage* is that it is possible to predict failure under service conditions with relatively small samples, and with no prior knowledge about the true mean or standard deviation of the distribution; (2) the *disadvantages* are that some distribution assumption of critical responses must be made which must be very exact if predictions far out on the tails are to be valid, and some knowledge of the 20% and 80% failure points must be had; (complete failure or complete success is worthless information for this test).

V. TESTS BELONGING TO CATEGORY I(d)

In the previous paragraph we have given, in essence, the treatment of sensitivity data as developed by Bliss. This has opened the way directly to a discussion of problems belonging to category I(d) (i.e., complete run-down tests where the stimuli used range from the stimulus for which none of the objects is affected to the stimulus for which all of the objects tested are affected). Bliss's method is applicable to such problems. It is, however, the purpose of the remaining portion of the paper to give an alternative method of analysis which we feel is to be preferred primarily because of its simplicity. The method was first proposed by Spearman in the treatment of problems arising in psychometrics and was recently generalized by the authors.[11] We will assume throughout this treatment that either the critical stimuli (or some simple function of the critical stimuli such as the logarithm) are normally distributed. This sort of presupposition seems to be in accordance with the findings of experimenters in the field of sensitivity data and is certainly needed if there is to be any meaning to the estimation of low percentage points.[10] However, it can be shown that the normality of the universe is not a necessary condition in order that the method yield unbiased estimates of μ and σ^2.

[10] The estimate of a low percentage point (.1% or .00003%) is based on the presupposition that the universe under study is normal or nearly normal "in the tails" of the distribution. This could only be "tested" empirically by taking a very large sample at extreme points; and even if one sample confirmed the normality hypothesis, we would *not have any great confidence* in subsequent samples. The alternatives are to give up any attempt to estimate a low percentage point, or else to make certain presuppositions as a basis for inquiry. The latter alternative is not only sound methodology, but is characteristic of all scientific inquiry; for fuller explanation, the reader is referred to C. W. Churchman, "Probability Theory," *Phil. of Science*, *12*, 147–173 (1945).

[11] C. Spearman, "The Method 'Right and Wrong Cases' (constant stimuli) without Gauss' Formula," *British Journal of Psychology*, *2*, 1908, 227–242 and B. Epstein and C. W. Churchman, "On the Statistics of Sensitivity Data," *Annals of Mathematical Statistics*, *15*, 90–96 (1944).

In brief, the attack on problems in category I(d) using Spearman's method is based on the key assumption that if p_i is the observed fraction unaffected at stimulus x_i and p_{i+1} the observed fraction unaffected at stimulus x_{i+1}, then $p_i - p_{i+1}$ is an estimate of the fraction just affected (i.e., the fraction of those that have "critical" responses) at about $\frac{1}{2}(x_i + x_{i+1})$. If we assume initially that the x_i are equally spaced and that $p_l = 1.0$ and $p_s = 0$, then the set of data on critical responses may be transformed into data where the x_i's are integers and the intervals are unity. In problems in the detonation of solid explosives it is found convenient to have the strength of blows increase linearly. In biological problems it is customary to choose the dosages in such a way that their logarithms are equally spaced. To take the x_i (or $\log x_i$) as equally spaced is convenient in practice, and this furthermore leads to a great simplification in the computations. However, it is by no means essential that this condition be true. Formulae can easily be derived for $\bar{x}$, s, the higher moments, and the sampling errors, even if the x_i are not evenly spaced.

In the second reference given in footnote (11), one will find a complete method for determining $\bar{x}$, s, the higher moments, and their sampling errors in terms of the π_i, the true fraction unaffected. In practice, of course, one rarely knows the π_i but does know the p_i, the observed values of the fractions unaffected at stimulus x_i. It is possible to give unbiased estimates for $\bar{x}$, s and the sampling variances of these unbiased estimates in terms of the p_i, but the proofs will not be given in this paper. An outline of the key formulae is given in Statistical Note No. 3.

The general remarks about this test are: (1) the *advantage* is that with a relatively small sample size, and simple computations, one is able to predict failures over a wide range of applications without presupposing the parameters (mean, standard deviation, etc.) of the distribution of critical responses; (2) the *disadvantage* is that for prediction purposes, the form of the distribution function of critical responses must be presupposed, especially for predictions of very low percentage points.

By making the transition from category I(a) to category I(d), a step of great importance has been taken. We have been able to increase enormously the efficiency and precision of the inferences that can be drawn from small samples. This is chiefly due to the fact that we are now able to treat the sensitivity problem in such a way that an attributes problem has been translated into one involving a continuous distribution. This means that even though we cannot measure the

critical stimuli for any of the objects under test, we can still associate with any lot actual numbers, namely, $\bar{x}$, the estimate of stimulus for which 50% of the objects will respond, s, the estimate of the standard deviation of the critical stimuli, and finally by proper choice of k, determine the stimulus for which any preassigned fraction of objects will respond or fail to respond. Knowledge of $\bar{x}$, s, $\bar{x} \pm ks$, and their errors will prove very useful in comparing the qualities of two or more lots. Furthermore, control charts on $\bar{x}$, s, and $\bar{x} \pm ks$ when applied to a succession of lots coming from a common source are very informative in detecting significant shifts in sensitivity level or the lack of control in individual lots.

It will be noted that the choice of the four tests discussed here depends upon two things: the purpose of the test and the amount one is willing to presuppose. If the purpose of the test is to predict over a wide range of possible applications, tests under I(a) are the least satisfactory. It is to be emphasized that presuppositions must be made in every test procedure and the strength of the presuppositions must be weighed against cost and the confidence based on previous experience.

VI. OPERATING CHARACTERISTICS OF TESTS OF INCREASED SEVERITY

The discussion contained in the previous sections can be summarized conveniently by operating characteristic curves (OC curves). These curves really tell the whole story behind an inspection plan or an experimental design for they give the probabilities of acceptance (or rejection) for any given "true" state. For the method of calculation of the OC curves discussed in this section see Statistical Note 4.

As a simple example under I(a), assume that the inspection plan demands acceptance only if no defects occur in a sample of 20 (i.e., reject if one or more defects occur in the sample.) The resulting operating characteristic curve is shown as curve A in Fig. 2; curve B represents an identical plan except that 100 are tested instead of 20. It will be noted that curve B is moved to the left of curve A, which means that the chance of accepting bad material is always less under this plan (as is to be expected, since the requirement that no defect occur in 100 is much harder to meet than the requirement that no defect occur in 20). It will also be noted that curve B rises more sharply than curve A; a steep slope in the OC curve is always a desirable property for it means that the risks of rejecting good lots, and the risks of accepting bad lots are both less (provided the curve is properly "located"). Thus, the real advantage of an increased sample is to be found in the sharper

slope of the OC curve, and if this slope can be made sharper by increased severity rather than increased sampling, a great advantage is gained thereby.

The gain in an increased severity test is shown in Fig. 2. Plan C consists of testing, in accordance with I(b), at a stimulus (X_0) where about 50% of the standard product fails. It is assumed that the standard deviation of critical stimuli is unity and is constant for all lots; it is further assumed that the critical stimuli are normally distributed. The standard lot in this case was assumed to be one which would give only .05% failures at the specification stimulus. It will be noted that despite the fact that Plan C employs the same sample size as Plan B, the slope of the OC curve is much sharper for C. The risk of rejecting good material is less (.023 vs. .05), and the risk of accepting bad material is greatly reduced (e.g., if a lot has 1% defectives, Plan C would practically never accept it, whereas Plan B would accept it about 40% of the time).

In a large number of instances, the manufacturer is anxious that no defectives at all be found in the lot; of course, no sampling plan could ever give absolute assurance of this quality, but it may be extremely desirable to have a very small risk of accepting material with, say, .001% defectives or more. Such items are usually "critical" in some sense; that is, if the item is defective, complete failure of a rather costly manufactured piece will occur. (An excellent example is the primer or detonator in ammunition because these usually initiate the propellants and, if they fail, the entire round may be wasted.) In these cases the cost of the item is usually small (though not necessarily), and the manufacturer is willing to increase the producer's risk (risk of rejecting good material) in order to decrease the consumer's risk to an absolute minimum.

In such cases, provided we have no accurate knowledge of the variation of the standard deviation, the schemes discussed in categories I(c) and I(d) are practically necessary. Suppose we want at least a 50-50 assurance that the lot contains no more than three defects in 10 million (.00003%). Plan D represented in Fig. 3 consists of a rundown test in accordance with I(d), or of a two-stimulus test in accordance with I(c). In order to make comparisons we assume the errors of estimate of the mean critical stimulus and of the standard deviation to be about $.1\sigma$; this is a reasonable assumption for the error of these statistics in tests run at intervals of about one standard deviation, fifty samples at a stimulus, or in most two-stimulus tests where 100 to 200 are tested at each stimulus. The average number tested in either case ranges from 200 to 400. We assume the distribution to be normal

and require that the quantity $\bar{x}+5s$ be less than the specification stimulus.[10] Since estimates of $\bar{x}+5s$ are themselves approximately normally distributed, we will accept 50% of the lots having a true value of $\bar{x}+5s$ at the specification stimulus. The OC curve in this case is plotted vs. the log fraction failing (as was done in Fig. 2.) Plan E represents a single stimulus sequential sampling plan which intersects the OC curve for Plan D at the .05 and .95 probability levels.[11] The enormous amount of testing that Plan E involves is shown in Fig. 4. While the testing for Plan D involves an average sample of from 300 to 350 for each lot, the sampling for Plan E might require *as many as* 100,000 items!

Another important advantage of tests of increased severity, an advantage whose importance it is difficult to over-emphasize, lies in another type of OC curve shown in Fig. 5, which cannot in general be constructed when one tests at only one stimulus. This OC curve can only be plotted when quantitative measurements are taken, or when, as in this case, qualitative judgments are transformed into quantitative.

The OC curve of Fig. 5 is derived from Plan D and plots the probability of acceptance against the stimulus ($X_{max.}$) where only three defectives in ten million occur. It will be seen that practically no lot will be accepted, the $X_{max.}$ of which exceeds the "specification stimulus" by more than 1.25 standard deviations.

This information is extremely important since it provides an exact basis for setting a "safety factor" for inspection tests; we always want the specification stimulus to be more severe than the conditions to which we expect the product to be subjected in service, but the question that always puzzles the careful specification writer is the amount of safety margin to introduce. For example, suppose we know that a certain weld will practically never receive a blow in excess of 150-foot pounds in practice. What value should be chosen for the blow which all pieces are supposed to pass? We ought not to choose 150-foot pounds, for any sampling plan leads to errors, and we require some margin to take care of these errors. What is often done is to select some arbitrary figure, such as 200 or 250-foot pounds, on a vague guess as to the adequacy of the safety margin. If an inspection plan similar to Plan D is used, such arbitrary guesses can be avoided. From tests on the product we can estimate the highest standard deviation to be expected and choose a value which exceeds 150-foot pounds by more than 1.25 standard deviations.

[11] See "Sequential Analysis of Statistical Data: Theory" and "Sequential Analysis of Statistical Data: Applications," *AMP Report* Nos. 30.1 and 30.2R.

There is a difficulty of a serious nature in all uses of arbitrary safety margins. An arbitrarily selected margin will either seriously increase the rejections of perfectly good lots, or seriously increase the acceptance of bad. Such arbitrary margins really operate in the worst possible manner: if the lot is excellent in quality, the margin is a handicap, whereas for bad lots, where "safety" is most required, the margin is apt to be dangerously small. Thus, the need of exact procedures, where they can be employed, is imperative.

The enormous saving in sample size of and the amount of information to be gained from an accurately designed increased severity test is really no more than an example of the saving introduced by any scientific advance. The saving indicates strongly the desirability of experimental investigations whose purpose is to provide adequate increased severity tests.

VII. ILLUSTRATIONS OF THE THEORY OF COMPLETE RUN-DOWN TESTS

We shall illustrate the theory treated in Statistical Note No. 3 by two examples. The first example will involve finding out whether or not a certain lot meets a specification requiring that only 1 in 1,000 objects fail to be affected at a certain preassigned stimulus. The second example will involve the comparison of two lots as to possible differences in their quality.

It will be recalled that in order to ensure distinguishing between a lot which has 1% failures and another which has .1% failures at the specification stimulus, it was necessary to test about 600 specimens. We also saw that sample sizes averaging about 20 specimens are adequate, *if the tests were conducted at the 50% increased severity level and if the assumption is made that the critical stimuli for all lots are normally distributed with the same standard deviation.*

We shall find that we are generally required to test considerably more than 20 and in general considerably fewer than 600 specimens when we use the complete run-down. Despite the fact that we use more than 20 specimens in our sample, our test is far more efficient than the increased severity test taken at one level alone since we obtain valuable information on the distribution of critical stimuli without making restrictive assumptions regarding the nature of the distribution.

Turning now to our first illustration of the theory contained in Note 3, we consider the following problem: an insecticide is being tested and it is desired to know whether or not the insecticide is so effective that all but 1 in 1,000 insects will fail to survive 40 milligrams of the insecticide.

The test was carried out by testing 20 insects at each of the following dosages: 5, 10, 15, 20, 25, 30, and 35 milligrams. The results were as follows:

Dosage	Number Tested	Number Surviving	Fraction Surviving
$x_1 = 5$ mg.	20	20	$p_1 = 1.0$
$x_2 = 10$ mg.	20	19	$p_2 = .95$
$x_3 = 15$ mg.	20	18	$p_3 = .90$
$x_4 = 20$ mg.	20	14	$p_4 = .70$
$x_5 = 25$ mg.	20	4	$p_5 = .20$
$x_6 = 30$ mg.	20	1	$p_6 = .05$
$x_7 = 35$ mg.	20	0	$p_7 = 0$

If we denote the dosages by (x_i), it is convenient to transform the set of dosages (x_i) into a set (x_i') where

$$x_i' = \frac{x_i - 5}{5}, \quad x_i' = (0, 1, 2, 3, \cdots, 6).$$

Noting that our formulae for the mean critical stimulus and the standard deviation of the critical stimuli involve all the p_i exclusive of $p_1 = 1.0$ and $p_7 = 0$, we find it convenient to tabulate the data as follows:

TABLE 2

x_i'	p_i	$b_{2,i-1}$	q_i	p_iq_i	$b_{2,i-1}p_i$	$b_{2,i-1} - 2\Sigma p_i$	$(b_{2,i-1} - 2\Sigma p_i)^2$	$(b_{2,i-1} - 2\Sigma p_i)^2 p_iq_i$
1	.95	1	.05	.048	.95	−4.6	21.16	1.02
2	.90	3	.10	.09	2.70	−2.6	6.76	.61
3	.70	5	.30	.21	3.50	− .6	.36	.08
4	.20	7	.80	.16	1.40	1.4	1.96	.31
5	.05	9	.95	.048	.45	3.4	11.56	.55

It will be found that:

$\sum p_i = 2.8$; $\sum b_{2,i-1}p_i = 9.0$; $\sum p_iq_i = .56$; $\sum (b_{2,i-1} - 2\bar{x}')^2 p_iq_i = 2.57$.

We are now in a position to substitute in the appropriate formulae in Statistical Note No. 3. The results are:

$$\bar{x}' = x_1' + .5 + \sum p_i = .5 + 2.8 = 3.3$$

$$s^2 = \sum b_{2,i-1}p_i - (\sum p_i)^2 + \frac{\sum p_iq_i}{n_i - 1} + G(\mu_2)$$

$$= 9.00 - 7.84 + .03 - .08 = 1.11$$

$$\therefore \; s = 1.05.$$

Also

$$s_{\bar{x}'}^2 = \frac{\sum p_i q_i}{19} = \frac{.56}{19} = .029$$

$$\therefore \quad s_{\bar{x}'} = .17$$

and

$$s_{s'}^2 = \frac{\sum (b_{2,i-1} - 2\bar{x}')^2 p_i q_i}{19} = .136$$

$$\therefore \quad s_{s'} = .37$$

but

$$s_s = \frac{s_{s'}}{2s} = \frac{.37}{2(1.05)} = \frac{.37}{2.10} = .18.$$

Transforming into the original x units it follows that the estimate of *the mean critical stimulus is equal to 21.50 milligrams. The estimate of* the standard deviation of the mean critical stimulus is given by (.17) (5) = .85 milligram. It can be shown that $\bar{x}$ is very nearly normally distributed even for small samples[11] and, therefore, at a 95% confidence level, the true mean critical stimulus (μ) should be within:

$$21.50 \pm (1.96)(.85) = 21.50 \pm 1.67 = 19.83 - 23.17 \text{ milligrams.}$$

At a 99% confidence level, μ should be within:

$$21.50 \pm (2.58)(.85) = 21.50 \pm 2.19 = 19.31 - 23.69 \text{ milligrams.}$$

The estimated standard deviation of the critical stimuli in original units is given by (1.05)(5.00) = 5.25 milligrams and the standard deviation of the standard deviation is equal to .90 milligram. Again it is possible to set confidence limits on s.

Let us recall, however, that we want fewer than 1 in 1,000 insects to survive a dose of 40 milligrams. In terms of the transformed x' units this means that the 99.9% point should not exceed 7.0. But, it is easily seen that $\bar{x}'+3.09s$ (the 99.9% point) $=3.3+(3.09)(1.05)=6.54$ with a standard error of:

$$s_{99.9}^2 = s_{\bar{x}'}^2 + (3.09)^2 s_s^2 = .028 + (9.6)(.0324) = .028 + .311 = .339$$

$$\therefore \quad s_{99.9} = .58.$$

The probability of getting a sample value of 6.54 from a universe

[11] The fact that $\bar{x}$ and s approach normality rapidly as a function of N, the number tested at each stimulus, has been established by the authors. The results have not yet been published.

possessing mean 7.0 and standard deviation .58 is found by noting that 6.54 is .79 standard deviation away from 7.0. There is approximately one chance in five of getting the 99.9% point to be 6.54 or less and we cannot assert that the insecticide meets the requirement at 40 milligrams on the 5% or even the 10% level of significance.

It is easy to give an example of the comparison of two lots. The calculations of the statistics of the lots would, of course, follow the same pattern as that described above. For example, suppose that the first lot had the characteristics just described and that the second lot possessed an average critical stimulus (in transformed units) $\bar{y}'=2.7$ with $s_{\bar{y}'}=.15$; $s_{y'}=1.3$, $s_s=.14$.

It is then readily seen that the average critical stimuli differ significantly for the two lots since

$$t=\frac{\bar{x}'-\bar{y}'}{\sqrt{s_{\bar{x}'}^2+s_{\bar{y}'}^2}}=\frac{3.3-2.7}{\sqrt{(.17)^2+(.15)^2}}=\frac{.6}{.22}=2.7.$$

The probability of getting a difference as large as .6 is less than 1 in 100 and therefore the lots differ significantly as to their mean critical stimuli. The standard deviations of the critical stimuli do not differ significantly. Computation will also show that the 99.9% points differ significantly on the 5% level of significance.

VIII. CONCLUSIONS AND PROPOSED INVESTIGATIONS

These examples are by no means intended to exhaust the possibilities of the application of the theory of complete run-down tests as applied to sensitivity data. In another paper it will be shown how one can make modifications of this approach to the case where the test is not a complete run-down and where the data are given over only a partially complete set of stimuli. It also will be shown that in some instances more precise inferences concerning $\bar{x}$, s, and $\bar{x}\pm ks$ can be made by varying the sample size tested at each stimulus. These problems are of key importance from a practical standpoint and merit consideration in a separate paper.

In conclusion, the authors wish to emphasize that it has not been their object to treat the subject of increased severity tests exhaustively. The key purpose of the paper is to bring the various methods into bold relief, to compare these methods in some of their aspects, and to point out some of the possibilities inherent in the newer methods. The authors hope that this paper will serve to stimulate the application of modern statistical methods to tests of increased severity by design and research engineers.

STATISTICAL NOTE 1

Stated in statistical terms, the problem is to estimate the universe average μ and the standard deviation σ (and also their standard errors) of the random variable x, which denotes the stimulus required to obtain a critical response of any individual drawn from some universe of objects. This problem differs from the usual one arising in frequency distributions since the critical stimulus cannot be measured for any individual in the universe. This forces us to find estimates of μ and σ knowing only the estimates p_i of the fraction of objects surviving at a discrete set of stimuli x_i, $i=1, 2, \cdots, n$. This approach is essentially the one followed in the "probit" method and other modern treatments of the dosage-mortality problem and tests of increased severity.

STATISTICAL NOTE 2

It can be shown using the first term of the binomial expansion that the chance $L(N, Q)$ of finding no defectives in a random sample of size N drawn from a lot with true fraction defective Q is given by $L(N, Q)=(1-Q)^N$ (where it is assumed that non-replacement of an inspected article does not affect the quality of the remainder of the lot). The results in Table I may be obtained by letting $N=20$ or $N=100$ and assigning various values to L. For example, if $N=100$ and $L(N, Q)=.5$, then $Q=.007$. This means that there is a 50–50 chance of accepting lots which are .7% defective, if acceptance is based on finding no defects in a sample of 100.

STATISTICAL NOTE 3

It is assumed in the following that p_i is an unbiased estimate of the fraction unaffected at stimulus x_i and that p_i-p_{i+1} is an unbiased estimate of the fraction unaffected at stimulus $x_i+x_{i+1}/2$. In practice it will be found convenient to choose intervals $x_{i+1}-x_i$ whose lengths are approximately given by σ. However, in some cases good results can be obtained by making the lengths of the intervals as large as 2σ. For convenience, we shall assume that the interval of test is unity: $x_{i+1}-x_i=1 (i=1, 2, \cdots, n-1)$.

Under the above restrictions we shall show how the average critical stimulus (denoted by μ), the standard deviation of the critical stimuli (denoted by σ), and the higher moments may be easily estimated. We shall also show how the sampling errors of the estimates $\hat{x}$, s and the higher moments can be found. A brief summary of the proofs and results will be given. These formulae depend for their validity upon certain assumptions regarding the interval of test and the distribution function of the critical stimuli.

If p_i denotes the observed fraction of objects unaffected at x_i and p_i' is defined as $p_i'=p_i-p_{i+1}$ then the key formulae are as follows:

The estimated mean about the origin is given by:

$$\hat{x} = \sum_{i=1}^{n} (p_i - p_{i+1}) \left(\frac{x_i + x_{i+1}}{2}\right) = \sum_{i=1}^{n} (p_i - p_{i+1})(x_i + 1/2); \tag{1}$$

the estimated qth moment about the origin is:

$$(2) \qquad m_q' = \sum_{i=1}^{n} (p_i - p_{i+1})(x_i + 1/2)^q,$$

and the estimated qth moment about the mean is:

$$(3) \qquad m_q = \sum_{i=1}^{n} (x_i + .5 - \bar{x})^q p_i .$$

If we take our assumed origin of coordinates at $x_1 + .5$, then it can be shown by simple algebraic manipulation that m_1' may be written simply as:

$$(4) \qquad m_1' = \sum_{i=2}^{n-1} p_i$$

where the summation is taken over all the p_i, except $p_1 = 1.0$ and $p_n = 0$. The value of $\bar{x}$ taken about 0 as origin is given by:

$$(5) \qquad \bar{x} = x_1 + .5 + \sum_{i=2}^{n-1} p_i$$

where x_1 represents the stimulus for which none of the samples will be affected. For the higher moments, it is easy to show that if m_q' is taken about $x_1 + .5$ as origin, then:

$$(6) \qquad m_q' = \sum_{i=2}^{n-1} b_{q,i-1} p_i$$

where $b_{q,i}$ represents the ith first difference of the consecutive qth powers of the positive integers, i.e., $b_{q,i} = i^q - (i-1)^q$.

For example, if $q = 2$,

$$b_{2,1} = 1^2 - 0^2 = 1; \qquad b_{2,2} = 2^2 - 1^2 = 3; \qquad b_{2,i} = i^2 - (i-1)^2 = 2i - 1.$$

In particular, $m_2' = p_2 + 3p_3 + 5p_4 + \cdots + (2n-5)p_{n-1}$.

Again, the origin of coordinates is taken at $x_1 + .5$.

Using the well known relationship connecting the qth moment about the mean with the moments about $x_1 + .5$ as origin,

$$(7) \qquad m_q = m_q' - q m_1' m_{q-1}' + \frac{q(q-1)}{2} (m_1')^2 m_{q-2}' + \cdots + (-1)^{q-1}(q-1)(m_1')^q,$$

it is possible to express m_q in terms of the p_i.
In particular for $q=2$, we have[11]:

$$(8) \qquad \begin{aligned} m_2 = s^2 = m_2' - (m_1')^2 &= p_1 + 3p_2 + 5p_3 + \cdots \\ &+ (2n-5)p_{n-2} - \left(\sum_{i=1}^{n-1} p_i\right)^2. \end{aligned}$$

It will be noted that we have been able to express the moments about the assumed origin $(x_1+.5)$ as linear functions of the p_i. As a direct consequence of this result, and also because the p_i are all independent, it follows that the standard errors of the various statistics are:

$$(9) \qquad s_{m_1'}^2 = \sum_{i=1}^{n-1} s_{p_i}^2 = \sum_{i=1}^{n-1} \frac{p_i q_i}{n_i - 1}.$$

$$(10) \qquad s_{m_q'}^2 = \sum_{i=1}^{n-1} b_{q,i-1}^2 s_{p_i}^2$$

where n_i = number of samples tested at the stimulus x_i, and

$$(11) \qquad s_{m_q}^2 = \sum (\beta_{q,i-1} - q\mu_{q-1})^2 s_{p_i}^2$$

where

$$(12) \qquad \beta_{q,i} = b_{q,i} - C_1^q b_{q-1,i}\bar{x} + C_2^q b_{q-2,i}\bar{x}^2 + \cdots + (-1)^q q \bar{x}^{q-1} b_{1,i}.$$

In particular in the case that $q=2$, it follows that:

$$(13) \qquad s_{m_2}^2 = s_{s^2}^2 = \sum_{i=1}^{n-1} (b_{2,i-1} - 2\bar{x})^2 s_{p_i}^2.$$

If (13) is combined with the well known result that

$$(14) \qquad s_s = s_{m_2}/2s, \quad \text{it follows that:}$$

$$(15) \qquad s_s = \frac{\sqrt{\sum_{i=1}^{n-1} \left[(2i-3) - 2\sum p_i\right]^2 s_{p_i}^2}}{2\sqrt{\sum_{i=1}^{n-1} (2i-3)\,p_i - \left(\sum p_i\right)^2}}.$$

[11] This is not, however, an unbiased estimate of σ^2. It can be shown that (6) is an unbiased estimate of μ and that $s^2 = \sum_{i=1}^{n-1} b_{2,i-1} p_i - \left(\sum_{i=1}^{n-1} p_i\right)^2 + \sum_{i=1}^{n-1} \frac{p_i q_i}{n_i - 1} + G(\mu_2)$, where $G(\mu_2)$ is Sheppard's grouping correction is an unbiased estimate of σ^2. Formulae (10) and (11) are the standard errors of these unbiased estimates.

It is often of interest to know not only the standard errors of $\bar{x}$ and s, but also to know the standard errors of $\bar{x} \pm ks$. Fortunately, the s^2 of $(\bar{x} \pm ks)$ can be expressed directly in terms of $s_{\bar{x}}^2$ and s_s^2 because of the fact that $\bar{x}$ and s generally have only a small correlation coefficient (<.1). Therefore, treating $\bar{x}$ and s as independent, it follows that:

$$\text{(16)} \qquad \text{var.}\,(\bar{x} \pm ks) = s_{\bar{x}}^2 + k^2 s_s^2 = \sum \frac{p_i q_i}{n_{i-1}} + k^2 \frac{\sum_{i=2}^{n-1} \left\{(2i-3) - 2\sum p_i\right\}^2 \frac{p_i q_i}{n_i - 1}}{4\left[\sum_{i=2}^{n-1} (2i-3)p_i - (\sum p_i)^2\right]}.$$

STATISTICAL NOTE 4

The *OC* curves for the various plans are computed as follows:

Plans A and B: For any example size n, the probability of acceptance (i.e., the probability of obtaining no defects) is P^n where P is the true fraction effective, i.e., $P = 1 - Q$, where Q is the value shown on the abscissa.

Plan C: In this plan the difference between X_1 (the specification stimulus) and X_0 (the increased severity stimulus) is 3.29 standard deviations since the standard lot is assumed to be one having .05% defectives at X_1. Bearing this in mind, it is easy to calculate (using the table of areas under the normal curve) the true fraction defective Q_0 at X_0 given Q_1 at X_1. For example, if $Q_1 = .0014$ for a certain lot then $Q_0 = .015$. The plan asserts that a lot is to be rejected if more than 00 defects occur in 100. The probability of accepting the lot with true quality $Q_0 = .015$ is formed by determining:

$$t = \frac{(.00 - Q_0)n^{1/2}}{[Q_0(1 - Q_0)]^{1/2}} = \frac{(.00 - .015)(100)^{1/2}}{[(.015)\ (.385)]^{1/2}} = -.31.$$

This means that the probability of acceptance is .38. It is to be noted that this method of computing the *OC* curve for Plan *C* is not generally applicable since t will not be normally distributed for all values of Q_0. When Q_0 is less than .10, the Poisson distribution represents an accurate estimate.

Plan D: Again Q_1 is given. Assuming a normal distribution of critical stimuli, we can solve for K in the equation:

$$\mu + K\sigma = x_1,$$

where μ and σ are the true mean and standard deviation, respectively, of the universe. The solution is obtained, as above, from tables of areas of the normal curve. A given lot will be acceptable if and only if $\bar{x} + 5s < x_1$, where $\bar{x}$ and s are the sample estimates of μ and σ, respectively. In order to determine the probability of acceptance, we have to determine the probability that we will estimate $\bar{x} + 5s$ to be less than or equal to x_1, when $\mu + K\sigma = x_1$. This is derived from:

$$T = \frac{x_1 - (\mu + 5\sigma)}{s_{\bar{x}+5s}} = \frac{\mu + K\sigma - \mu - 5\sigma}{s_{\bar{x}+5s}} = \frac{(K - 5)\sigma}{s_{\bar{x}+5s}}.$$

For the sample size tested, the estimates of μ and σ are approximately (and asymptotically) normally distributed, and hence so are the estimates of $\mu+5\sigma$, and T, if $\sigma_{\bar{x}+5s}$ is known and is constant. We assume $\sigma_{\bar{x}}=\sigma_s=.1\sigma$, and since $\bar{x}$ and s are approximately independently distributed, $\sigma_{\bar{x}+5s}=\sqrt{.26\sigma^2}$. Hence,

$$T=\frac{(K-5)\sigma}{\sqrt{.26}\,\sigma}=\frac{K-5}{.515}.$$

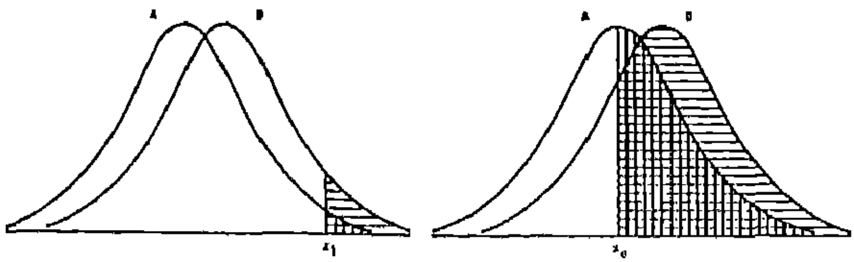
T can then be computed for any value of K, i.e., for any value of Q_1, and the required probability of acceptance can be determined from normal tables.

The OC curve for Plan D, showing the probability of acceptance vs. the stimulus ($X_{\text{max.}}$) where only three defectives in 10 million occur, is easily derived from:

$$T=\frac{X_1-X_{\text{max.}}}{s_{\bar{x}+5s}},$$

the same procedure being used as was used above.

FIGURE 1



x_1 = Specification Stimulus

Area under Curve "A" from x_1 out to + ∞ (area of defectives) = .1% of total area

Area under Curve "B" from x_1 out to + ∞ (area of defectives) = 1% of total area

x_0 = Increased Severity Stimulus

Area under Curve "A" from x_0 out to + ∞ = 50% of total area

Area under Curve "B" from x_0 out to + ∞ = 16% of total area

FIGURE 2

Operating Characteristic Curves for Single Stimulus Test
Without Increased Severity (Plans "A" and "B")
With Increased Severity (Plan "C")

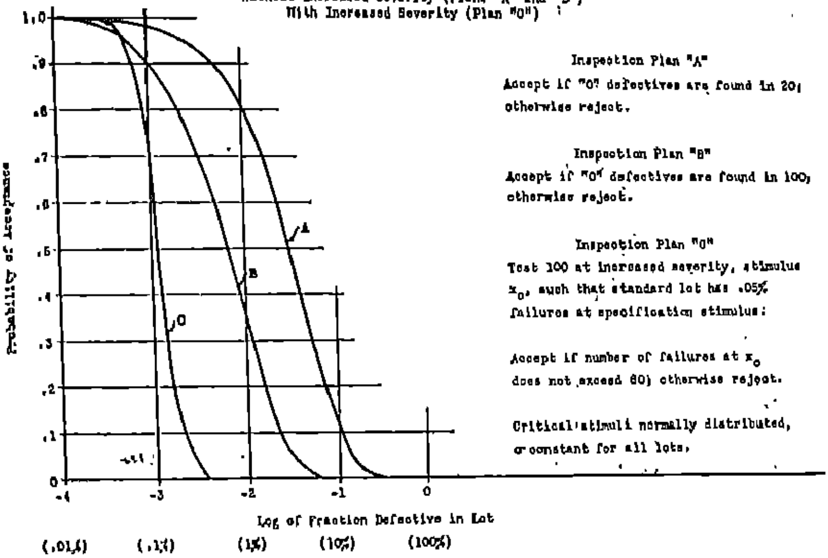


FIGURE 3

Operating Characteristic Curves for Run-Down Test (Plan "D")
and Best Single Stimulus Test (Plan "E")

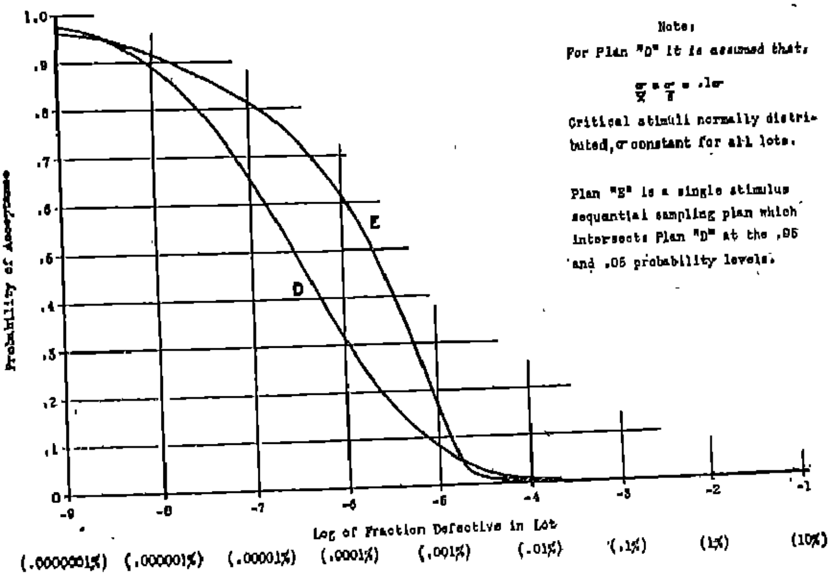


FIGURE 4

Average Sample Size Curve for Single Stimulus Sequential Test which Intersects Plan "D" at the .95 and .05 Probability Levels

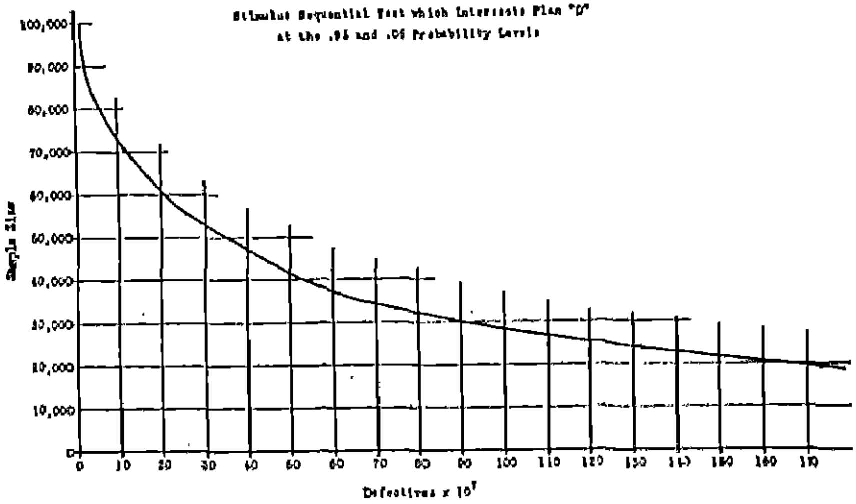
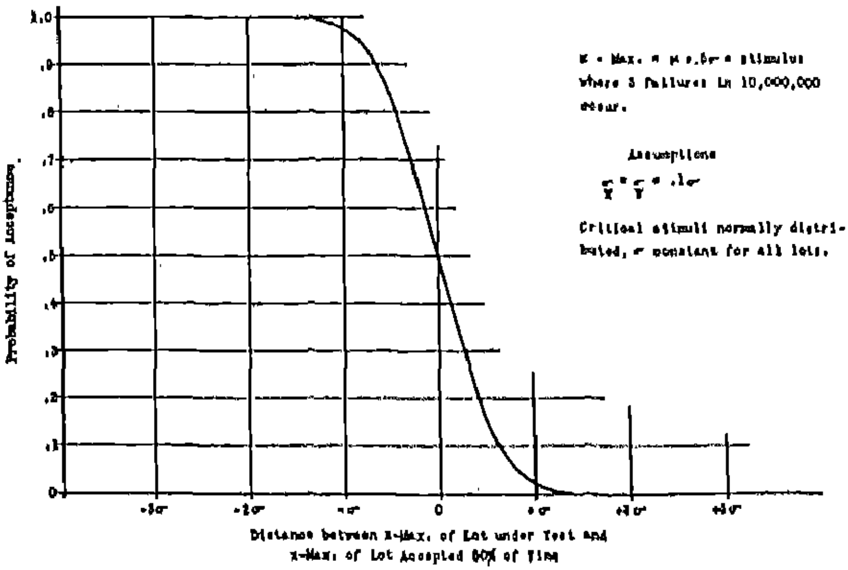


FIGURE 5

Operating Characteristic Curve for Run-Down Test
Probability of Acceptance vs. X - Max.



BOOK REVIEWS

Edited by
OSCAR KRISEN BUROS
Rutgers University

Instalment Mathematics Handbook: With Working Formulas for All Types of Transactions. *Milan V. Ayres* (Consulting Statistician, 5401 Woodlawn Ave., Chicago 15, Ill.). New York 10: Ronald Press (15 East 26th St.), 1946. Pp. xvi, 267. $10.00.

REVIEW BY CARL H. FISCHER
Associate Professor of Mathematics, University of Michigan

As its title implies, this book is intended to furnish "comprehensive and detailed information which will enable the user readily to make reliable time-payment calculations." The first of two parts consists of some 121 formulas and 13 pages of numerical tables. This section contains practically no explanation except for some illustrative examples. It is doubtful whether the class of users for whom this book was written, that is, "anyone familiar merely with arithmetic, simple algebra, and enough plane geometry . . . " could make use of this portion of the book without having previously gone carefully through the second section.

The second part, entitled "Derivations, Discussion and Proofs," constitutes the major portion of the book. Here, most of the formulas listed in Part I are developed, usually in considerable detail, and illustrated by numerical examples. The author has evidently had a great deal of practical experience in the field. This section of the book is liberally sprinkled with references to various practices, apparently common in instalment financing, some of which would certainly tend to discourage the lay reader from "buying on time."

There are three principal types of problems considered. The first is an internal analysis of finance company operations. Formulas are developed for such items as the rate of accumulation and of liquidation of instalment paper, an analysis of retail receivables, the collection ratio and life of notes, the ratio of borrowings to capital, and the distribution of earnings. These formulas, in the main, are based upon the arithmetic progression. Much of the analysis is quite ingenious although some parts of the exposition are not too clear, particularly to one not well versed in the instalment loan business.

The second type of problem given consideration is the determination of the size of the monthly payments under various kinds of common instalment contracts and interest assumptions. Here one is introduced to the meaning of such trade jargon as "balloon note contracts," "hold-back deals," "Morris Plan," "dealers' packs," and others. The chief criticism of the multitude

of formulas developed is that most are based upon the *uniform method* of determining yield, discussed below. From a practical standpoint, however, the formulas are undoubtedly satisfactory.

The third sort of problem is the determination of the yield to the finance company under various kinds of contracts. Six different methods are discussed, four of the "summation" type and two of the exponential. The author prefers the *uniform method*, an approximation, under which it is assumed that each instalment payment is apportioned in exactly the same way between reduction of principal and payment of charge or interest. The only justification for this procedure is the practical one of simplicity. Throughout the entire book, with stated exceptions, this is the interest method used. Another simple approximation rule, which is discussed only briefly, appears to have more theoretical merit. This is the *pro-rata method*, based upon the assumption that the portion of each payment devoted to the charge is proportional to the "total outstandings" at the start of that period, that is, to the sum of the outstanding principal plus the unpaid portion of the total charge. This deviates only slightly from ordinary compound interest.

It could be pointed out that one of the other summation processes discussed, the *residuary method*, under which it is assumed that all of the early payments are applied to reduction of principal and the charge is paid off last, produces a yield identical with that obtained by the use of the well-known "Merchants' Rule," under which the entire principal draws simple interest for the entire period; likewise, instalment payments made draw simple interest from the date of payment to the end of the period.

The author is aware that an exponential method is generally considered to be the preferred interest procedure but contends, somewhat surprisingly, that it is impractical for use in the ordinary finance company office because of its complication and because of the unavailability of the necessary interest tables. He seems to feel that there are two distinct exponential methods, labels one the *small loan* and the other the *present worth* method, and devotes several pages to an argument endeavoring to show the superiority of the latter. Analysis shows, however, that the two methods are identical except for the fact that, whereas the small loan method furnishes a nominal interest rate, converted monthly, the present worth method goes a step farther and produces the equivalent effective rate.

The author seems to overlook a similar point in chapter 18 where he goes into considerable detail in comparing two supposedly unlike methods for computing the yield on "hold-back deals." It is easy to show that his so-called "direct" method is merely a special case of the "reserve" method, under which the reserve fund earns interest at the yield rate.

There is rather an extensive chapter, the last one in the book, on interpolation. After a brief discussion of the linear case, the bulk of the chapter is devoted to the development of an interpolation method based upon fit-

ting certain hyperbolas to three given points and to an attempt to show the superiority of this method over the standard finite difference methods. As might be expected, the result is decidedly unconvincing; this chapter, which might well have been listed as an appendix, is easily the poorest portion of the book.

Such defects as have been mentioned here do not detract seriously from the importance of this book in its field. It should be indispensable in any finance company office, particularly if the state of mathematical erudition there is as implied by the author. In addition, the book should be considered almost equally valuable to teachers of the mathematics of finance and to authors of textbooks in the subject. After reading this handbook, the reviewer is impressed with the inadequacy of the treatment of instalment financing in the leading texts in the mathematics of finance. Even authors of textbooks in college algebra might find in this handbook some interesting and practical applications of arithmetic progressions with which they could enliven the topic, replacing the ever-present problems on the number of logs in a pile and the distance run in a potato race.

Mathematical Tables and Other Aids to Computation. Published quarterly beginning January 1943 by the National Research Council. (Number 15, July 1946 was the last issue available at the time of writing this review.) Beginning with 1947 the subscription price for each calendar year is $4.00, payable in advance; ordinary single numbers, $1.25. Earlier ordinary single numbers, each $1.00, and all numbers for each of the years 1943 to 1946 inclusive, $3.00. Special single number 7, Guide to Tables of Bessel Functions, $1.75, and number 12, $1.50. All payments are to be made to National Academy of Sciences, 2101 Constitution Avenue, Washington, D. C. Edited on behalf of the Committee on Mathematical Tables and Other Aids to Computation by *Raymond Clare Archibald* (Professor Emeritus, Brown University) and *Derrick Henry Lehmer* with the cooperation of *Leslie John Comrie* and *Solomon Achillovich Joffe.*

Review by Kenneth J. Arnold
Assistant Professor of Mathematics, University of Wisconsin

Statisticians use a wide variety of mathematical tables and computational procedures, some of which have been developed in statistics and others of which have been developed in astronomy, psychology, business, engineering, and various other fields. Even such statistical tables as distribution functions, percentage points, and constants of sampling distributions sometimes appear as appendices to articles devoted to particular applications and in journals not usually perused by statisticians. Frequently the general utility of such tables is disguised by headings in terms of the particular application as "No. of animals which die" for the upper limit in a binomial summation. Collections of tables such as Pearson's *Tables for Statisticians and Biometricians* and Fisher and Yates' *Statistical Tables for Biological, Agricultural and Medical Research* cannot obviate the necessity of reference to journals for tables needed in the solution of some of the

most frequently recurring problems. The continual appearance of new tables and new computational methods and the consequent increase in the problem of finding whether and where appropriate tables and methods are available are phenomena common to all fields in which computers and table users work. A journal devoted to the problem became a necessity and in January 1943 the first issue of *Mathematical Tables and Other Aids to Computation* appeared. The role of MTAC was described by Professor Archibald in the opening sentences of this first issue:

> This Quarterly Journal, a new publication of the National Research Council, is to serve as a clearing-house for information concerning mathematical tables and other aids to computation. Especially during the past decade have tools for computation been vastly multiplied. These tools, or accounts of them, are to be found in an enormous international range of book, pamphlet, and periodical publication, not only in the fields of Pure Mathematics, Physics, Statistics, Astronomy, and Navigation, but also in such fields as Chemistry, Engineering, Geodesy, Geology, Physiology, Economics, and Psychology. An attempt will here be made to guide varied types of inquirers to such material.

A summary of the contents of the first 15 numbers gives an assuring picture of the manner in which MTAC is fulfilling its purpose. The numbers 1 to 12, constituting Volume I, total 480 pages; the numbers 13 to 15 total 148 pages. Each ordinary number contains an article or articles on tables, computing machines, or computational methods. These articles are followed by sections entitled "Recent Mathematical Tables," "Mathematical Tables—Errata," "Unpublished Mathematical Tables," "Mechanical Aids to Computation," "Notes," "Queries—Replies," and "Corrigenda et Addenda." Special number 7 is a 104-page "Guide to Tables of Bessel Functions" by H. Bateman and R. C. Archibald. Special number 12 contains, in addition to the contents of an ordinary number, an index to Volume I.

Among the articles in numbers 1 to 6 and 8 to 15 are "Machines for Solving Algebraic Equations" by J. S. Frame and "Scientific Computing in Great Britain" by J. R. Womersley.

Under "Recent Mathematical Tables," 241 reviews have appeared. With each is associated a code letter or letters indicating the class of function or the field of application; as examples: *A* for arithmetical tables and constants, *C* for logarithms, *H* for numerical solution of equations, *I* for finite differences and interpolation, *K* for statistics, *N* for interest and investment, *O* for actuarial science, *Z* for calculating machines and mechanical computation. Of the 241 reviews, 23 have associated *K*'s. Each review contains definitions of the functions tabulated, the range of arguments, the number of decimals or significant figures in the entries. Most reviews also discuss accuracy, errata, related tables and applications. The thoroughness of the reviews is indicated by the fact that W. G. Cochran's review of the second edition of Fisher and Yates' *Statistical Tables* occupies 4 pages, roughly equivalent to 6 pages in this review section.

Under "Mathematical Tables—Errata," 86 lists have appeared. There is no rejoinder to the comment made by R. C. Archibald in presenting errata

in several statistics books: "Authors of frequently used works in the field of Statistics display some carelessness in the preparation of tables they publish."

Under "Unpublished Mathematical Tables," 40 entries have appeared. Some of these tables have thus been available several years before publication; others were never intended for publication but have been used in the construction of published tables. Under the next four headings the following numbers of entries have appeared: "Mechanical Aids to Computation," 22; "Notes," 58; "Queries," 18; "Queries—Replies," 24.

Cross references are quite adequate. In each entry, reference is made to earlier or concurrent related entries. In each section, reference is made to material appropriate to that section contained in entries in other sections; for example, the reviews of recent tables which contain lists of errata are mentioned under the heading "Mathematical Tables—Errata." The following two adverse criticisms will probably be invalidated in a short time. The over-long delay in the publication of reviews of some recent statistical tables can perhaps to a large extent be attributed to the delinquency of reviewers. While the Index to Volume I is inadequate as an aid to discovering tables of a particular function, Professor Archibald informs me that a detailed "Subject Index" covering the whole of Volumes I and II is planned for inclusion in the Index to Volume II which will probably appear in October 1947.

It can be hoped that a guide to mathematical statistical tables will appear in the not too distant future but a guide cannot replace the listing of errata and the careful reviews of MTAC.

Industrial Experimentation [Revised Edition]. *K. A. Brownlee* (Research Department, Distillers Co., Ltd., Great Burgh, Epsom, Surrey, England). A revision of the 1945 memorandum of same title reviewed in this JOURNAL, March 1946. Directorate of Ordnance Factories (Explosives), Ministry of Supply. London, W.C. 2: H. M. Stationery Office, 1946. Pp. 110. 2s. Paper. (New York 20: British Information Services (30 Rockefeller Plaza), 1946. $0.60.) *Two reviews follow:*

REVIEW BY GEORGE W. BROWN
Research Associate Professor
Statistical Laboratory, Iowa State College

THE revised edition is generally of the same character as, and in most respects identical with, the first edition. It has, however, been somewhat expanded in certain respects, notably with respect to multiple correlation and analysis of variance. Other additions, mostly in the form of explanatory remarks and cautions, contribute to a general improvement.

This reviewer is in substantial agreement with most of the comments, both favorable and adverse, made by Tukey and Wolfowitz in their reviews of the first edition. In fairness to the author, it should be pointed out that the present edition had passed galley proof stage before Mr. Brownlee had an opportunity to read the reviews in question.

This manual is, to use Professor Tukey's language, "a very good cookbook" of statistical methodology, and could, with more careful attention devoted to the basic hypotheses underlying statistical methodology, evolve into a manual of considerable stature in the field.

The major change in this edition consists of a reshuffling and expansion of the analysis of variance discussion, with greater attention devoted to background material and consideration of examples of multi-factor experiments with various orders of replication. Among the miscellaneous additions are a short section on the Poisson distribution, more introductory material on quality control, and inclusion of the range-standard deviation conversion, the Fisher inversion procedure in multiple regression, and properly discouraging remarks with respect to extrapolation outside the range of independent variables in regression work.

A few important corrections have been made, including a note on the two-tailed F test and a paragraph designed to remove the false impression that multiple correlation technique does not permit the study of interactions. In discussing an example of the significance of a mean, the author points out that sigma is an estimate; hence, the determination made on few degrees of freedom is less accurate than the determination made on a large number of degrees of freedom; he points out, moreover, that the "t" distribution takes this into account. This point, not made in the first edition, may remove some of the confusion resulting from the apparent identification of population variance and its sample estimate.

An added remark on page 20 seems to imply that, in testing homogeneity of variances, if the extreme estimates did not differ significantly by F, then Bartlett's test would not show significance. It seems to this reviewer that the remark in question should have been accompanied by a certain amount of caution.

The recommended procedure, given on pages 21 to 23, for comparing the means of two samples, depends on the sum of the sample sizes. It is not clear whether the author realizes that the following procedures are really different, but he suggests, without clarification, that if $n_1+n_2<30$ the variances be treated as homogeneous; that the Behrens-Fisher test be used when $n_1+n_2>30$. The important point is that from the very outset no mention is ever made of the importance of background hypotheses and parallel information.

A multiple regression example presented in chapter 10 has independent variables x_w, x_a, x_s. The regression on x_s alone is significant at the one percent level, but the effect of x_s is not significant when x_w and x_a are included. In discussing this situation on page 63, Mr. Brownlee says, in referring to the regression on x_s alone, "x_s would appear significant at the 1% level, an entirely erroneous conclusion." Why call this erroneous? The author is certainly aware that the regression on x_a and/or x_w might turn out not significant when adjoined to x_s. In any case, the very important question of the order of elimination of independent variables, or the corresponding question in

analysis of variance problems, seems nowhere to be mentioned. This is not a question of mathematics, it is an operational question whose solution is dictated by consideration of the problem at hand. Similarly, when two independent variables are correlated, the author says that the individual regression coefficients are fictitiously large. This would seem to imply that it would be incorrect to use an individual regression coefficient if either one of the independent variables is used alone. Again, the reviewer is certain that what was meant was that if two independent variables are to be used jointly, the regression analysis should be the joint multiple analysis.

On page 107, it is implied that the analysis of variance is impossible when the data are haphazard or when units are missing. This is not strictly true, although the analysis in such cases may be extremely difficult because of the large number of linear simultaneous equations which may have to be solved. On page 108, it is pointed out that the computation for a multiple correlation, with say 5 independent variables, is much more severe than for a five factor analysis of variance. It might be noted that the analysis of the conventional design has been simplified by the introduction of various orthogonalities; similar orthogonalities in a multiple correlation analysis would simplify the analysis correspondingly. When hypotheses are properly formulated in the analysis of variance situation and in the multiple correlation situation, the computations coincide as well as the significance tests. The major difference between the two situations is primarily with respect to the way parametric hypotheses are stated.

On page 31, it is remarked, in connection with 2×2 contingency tables, that in borderline cases the correction for continuity is important. Since the chi-square distribution is still an approximation, in this case, with or without the correction for continuity, and since the statistician rarely has adequate quantitative basis to set precise significance levels, it would appear to this reviewer that no really important decision should rest on the correction for continuity.

This reviewer feels that the value of this pamphlet could still be increased by further revision to include expanded background exposition of the hypotheses behind the statistical tests, discussion of "practical" significance and perhaps points such as those raised above, where further clarification might be useful as well as informative. Nevertheless, the pamphlet in its present form is potentially useful to a large class of people and contains a considerable amount of material for the rather low price of sixty cents.

REVIEW BY ALAN E. TRELOAR
Associate Professor of Biostatistics, University of Minnesota

THIS interesting manual of statistical methods, written primarily as a guide for industrial research workers who are not acquainted with the procedures or philosophy of statistics, has been considerably extended in the present revision. It is now referred to in the preface as a "monograph" instead of a "report," and the author's name is appropriately displayed on the

title page instead of being given merely in an acknowledgment line of the foreword as was previously the case.

The textual material of the earlier edition has been amended somewhat through incorporation of some additional explanatory sentences, alternative procedures and extra examples, all of which are valuable. However, the principal changes involve (1) reorganization and extension of the discussion of analysis of variance in relation to factorial experimentation, (2) illustration of logarithmic transformation to reduce abnormality and permit routine handling of proportional change situations, and (3) inclusion of a short chapter on the Poisson distribution. The 27 extra pages of text, together with removal of line spacers at many points (without much loss in reading ease), have increased the volume of discussion by perhaps 40 per cent.

The author had not seen the reviews of the first edition given in this Journal when the revision went to press. In a communication to the Review Editor he refers to this as "unfortunate." The present commentator agrees that this is so; for had adjustments been made to eliminate the defects discussed in those excellent reviews, the present edition would have had considerably enhanced value.

The error of using the F test when $|F|$ is called for has been corrected in this revision, although the approach is somewhat clumsy because reference to the correct answer is given as an afterthought to use of the inappropriate one. It is also stated without reservation that the probability corresponding to any given $|F|$ is double that for F, but this is true only when the samples are of the same frequency. The other defects discussed by Professors Tukey and Wolfowitz remain. It may be added that correction for continuity is considered of great importance in 2×2 tables but the subject is not mentioned in connection with a 3×1 table where each of the expected frequencies is only 5. Also a reversed code scale in the example of simple correlation may cause trouble for the inexperienced reader, for the sign of the coefficient is reversed without explanation.

The reiterated conclusion that an insignificant F means that an initially assumed component of the numerator variance does not exist is most disturbing to this reviewer. Perhaps the author is uneasy about this point when he concedes that values of F somewhat below the accepted critical level for a claim of significance suggest situations worthy of further investigation. However, he follows the beaten path in usage of the 5 per cent level rather assiduously in the actual analyses.

On page 13 one finds the following statement, which is not uncommon in statistical texts: "For many purposes the 5 per cent level is accepted, but we must realize that this means that 1 in every 20 times we will assert that an effect exists when it really does not." The assumption implicit here that all responses are in fact expressions of chance variation is a strange one, and quite out of line with much of the argument in this book. The *maximum* proportion of errors "of the first kind" corresponds to the significance level adhered to; usually, the proportion may be expected to be very much less.

While the author seems aware of the dangers of assuming that actual data

conform in structure to the mathematical models in terms of which the statistical procedures have been elaborated, he does not discuss the assumptions intrinsic to the models at all adequately. Indeed at one point (p. 108) he goes so far as to say: "we can disregard the underlying assumptions [of the analysis of variance] with comparative impunity." Many will challenge this point.

This book will prove useful to teachers of statistics, despite its failure to maintain desirable technical standards in the respects mentioned. It broadens the horizon with applications in a field where before there was but little illustrative material. Although the theme is pursued with missionary zeal, unusually good balance with practical considerations is preserved. Provided that the reader is aware of the deficiencies discussed in these reviews, the presentation should prove very helpful as accessory reading in connection with statistical courses.

Measuring Business Cycles. *Arthur F. Burns* (Director of Research; Professor of Economics) and *Wesley C. Mitchell* (Director of Research, 1920–1946; Professor Emeritus of Economics). (National Bureau of Economic Research; Columbia University). Studies in Business Cycles, No. 2. New York 23: National Bureau of Economic Research, Inc. (1819 Broadway), 1946. Pp. xxvii, 560. $5.00.

REVIEW BY ELMER C. BRATT
Professor of Economics, Lehigh University

THE publication of Number 2 of the National Bureau's Studies in Business Cycles is a notable event. This volume is faithful to the general pattern of study outlined in 1927 when Number 1 was published but has appeared later than then anticipated. In the meantime the program as a whole has taken shape. Number 2 highlights the methods employed and evaluates their effectiveness. It is to be followed by a series of monographs applying the techniques here established to separate economic processes. A final volume is projected to weave the results together into a theoretical account on how business cycles run their course, although a summary preview is now promised for the near future.

The methods of analyzing the business cycle are developed to test the business-cycle hypothesis: recurrent expansions and contractions occurring at about the same time in many economic activities. The first step is to adjust seasonally the original series. Turning points are then located by judicious employment of a set of mechanical rules, establishing separate cycles in each series. Each of these cycles is then represented in percentage variations from its average value. "Specific Cycles" are thus obtained, correction having been made for seasonal and inter-cycle trend.

The specific cycles contrast with the "Reference Cycles," computed similarly for each series. The reference cycles cover the periods of general business cycles. Turning points in general business cycles represent the consensus of a collection of time series rather than points in a single aggregate index. Both the specific and reference cycles, where the data are monthly, are divided into nine stages, including the initial and terminal trough, the peak,

and three approximately equal periods each for expansion and contraction. The levels at the troughs and peak ordinarily are represented by a three month average. A series of tables for each series give a wide range of information, including the levels and duration of various stages, rates of change between stages, and inter-cycle changes. One set of tables is made for specific cycles and another for reference cycles. An additional set of tables is developed to show the conformity of each series to general business cycles. In each work table, values are given for the separate cycles and for the arithmetic mean and average deviation. By mid-1942 over a thousand monthly and quarterly series had been so analyzed, principally drawn from the United States, but over 200 from the three principal European countries. In addition, over 200 yearly series have been analyzed, a third of them from Europe where monthly data are less available. The yearly data are found to be a poor substitute, however, because major cycle changes occur within a year.

But little of this work-table information could be presented in the present volume. After providing a full description of the methods, attention is turned to issues regarding the effectiveness of them. First, consideration is given to the retention of the intra-cycle trend. Conviction here resulted from the fact that the trend line determined depends upon subjective factors together with a belief that the trend must be retained to represent the "cycle of experience" as the unit of analysis. Trend adjustment in six series is found to reduce cyclical variability and makes the cycles in different processes more comparable. Supplementary trend-adjusted data are not employed for all series simply because of the work and expense.

The need for smoothing out random influences initially is checked by contrasting results obtained from smoothing four actual and six artificial series, principally by Macaulay's "43 term summation approximately fifth-degree parabolic graduation." It is concluded that initial smoothing would eliminate part of the actual cyclical movement but not all of the random influence.

Major attention is given to the significance of the average cycles computed. Since cycles are thought of as unique experiences by the National Bureau, the employment of averages presents difficult problems. Should cycles be grouped secularly or according to some scheme of long cycles before averaging? Little relationship is found between order in time and duration and amplitude of 7 test series; secular changes in durations and in amplitudes do not appear to be significant. Checks on the variation in character of business cycles presupposed by the most acceptable long-cycle hypotheses reveal no significant grouping.

Measurements presented in the volume move in almost all directions. No measurement is considered final; all are viewed as steps in a series of successive approximations. The American monthly reference dates through 1927, for instance, have not been revised since 1929, although some are definitely known to require revision. The conclusions cannot be applied as generally as many of us would like; they apply rather narrowly to the significance of the National Bureau methods.

Do the benchmarks set out provide appropriate background for the forthcoming analyses of various economic processes? The arithmetic mean may be called into question. Although the authors are convinced that the median and positional means are less satisfactory, the possibility of the geometric mean is not considered. The prevailing positive skewness of cycle-length distributions appears to the reviewer to involve more than the artificiality introduced by setting a minimum but no effective maximum length for the cycle.

The rejection of positional means in averaging cycles and acceptance of them in measuring seasonals may indicate that greater importance is attached to extreme values in the former case. Separate averages for major and minor cycles certainly would provide better clusters, and the reviewer believes greater homogeneity. Since the series employed do not represent processes, the amplitudes studied are of doubtful significance. Recognizing the inadequacies of all measurable amplitudes, the reviewer believes nevertheless that there is a clear difference in amplitude of industrial activity since 1878 between commonly recognized major cycles and others. The authors' data on length of expansions of deflated clearings show only three of seven rising from major depressions lasting less than thirty months and only three of eight rising from minor depressions lasting more than thirty months. Further, the authors note "the degree of cyclical diffusion is correlated with the amplitude of cyclical fluctuations."

Most of the cycle measures would be modified by displacement of the turning points employed. If the rules employed for determining these turning points are accepted, the question should be raised whether the measures developed overemphasize them. Although the turning points are dated at the end of a flat top or bottom, there are many cases where the date is several months before an expansion or contraction was recognized at the time. Since most series show interruptions of expansions and contractions while they are under way, why does the terminal high or low point have major significance? Notably, the timing measures, which are developed from a comparison of these turning points instead of from the unsatisfactory, but widely used, correlation method, are critically dependent upon this dating scheme.

Does the holding that the cycle of experience includes the intra-cycle trend, but excludes seasonal, represent a Schumpeter position to the effect that the trend and cycle are inseparable? Or does it result from statistical convenience? Many analysts will question the authors' position that the business community fully allows for seasonal and not at all for growth and decadence. It is one thing to recognize the difficulties of trend adjustment and another to rationalize that the trend is a part of the realistic cyclical experience. Does not the latter position prejudge the case to an extent that the character of cycle to be ultimately determined is partly postulated at the start?

The reviewer believes that per-month averages are overused. Which is a worse depression: one in which the contraction shows a violent and rapid

decline and therefore a large per-month average drop or one which shows a slower decline to the same depth?

The reviewer agrees that the frequent relaxation of rules, the making of separate judgments at successive stages of inelegant analyses, does not in this case represent a method inferior to more objective formulations. The degree of elegance we can afford depends upon the kind of problem faced. Careful study of the volume will give the student added respect for the importance of the business cycle as a distinct type of fluctuation.

Twenty years ago Mitchell's analysis was to be compared only with a multitude of competing theories, none of which had been adequately checked with measurable occurrence. The National Bureau methods are less preeminent today. Econometric analysis has proceeded a long way and has a substantial following. Concomitant-cycle analysis has a wide group of adherents. And the present renown of the Keynesian aggregative consumption-income theory puts it out of reach of earlier theories. In the present volume the National Bureau faces only the second of these contending frameworks, and in this case from the viewpoint of whether the bench mark averages are distorted by ignoring long cycles. Respect for the work of econometricists employing annual data is weakened, however, by the demonstrated inadequacy of annual data in studying business cycles. Although the volume does not materially reduce the area of disagreement, it represents only the initial evidence. We cannot expect the National Bureau to come to grips with competing frameworks until the processes are examined.

How to Read Statistics. *R. L. C. Butsch* (Formerly Professor of Education, Marquette University). Milwaukee 1, Wis.: Bruce Publishing Co. (524–544 N. Milwaukee St.), 1946. Pp. v, 184. $2.50. *Two reviews follow:*

REVIEW BY LOUIS GUTTMAN
Associate Professor of Sociology, Cornell University

AN EASY path to climb towards a serviceable knowledge of statistical theory is perennially being sought. The book now being proferred by Dr. Butsch is another attempt to convey statistical understanding to persons without any mathematical background. He addresses himself to the "large number of teachers, social workers, personnel directors, and industrial executives, who through lack of opportunity or inclination have failed to acquire the minimum of statistical techniques," that is, who have not ever been trained to apply a formula even in a purely mechanical fashion. The emphasis is to be on the "why" and not on the "how"; not a single formula appears in the book (albeit a few standard algebraic symbols are used).

The subject matter covered ranges from graphic presentation and averages through partial and multiple correlation, sampling errors, and analysis of variance and covariance. The examples used are from educational psychology.

The presentation is simple and lucid, on the whole. Indeed, it is one of the best-written statistical books this reviewer has seen. An unresolved problem in this reviewer's mind is: How long will the material remain with the reader?

Can a person sufficiently retain and appreciate the ideas that are laid out before him if he is given no practice in computing, not to speak of deriving the formulas that are the consequences of the ideas? This is a problem that well might be investigated experimentally by educational psychologists. In any event, Dr. Butsch's book can certainly provide a useful supplementary discussion to a more complete treatment, no matter how sufficient it is by itself for its intended audience.

A number of faulty statements that should be pointed out include the following: The author states that statistical treatment is required by sciences dealing with living things, but not by the physical sciences (p. 2); that dispersions of distributions cannot be compared unless they have the same average (p. 28); that standard scores are more "strictly mathematical" than percentile ranks (p. 73); that a correlation ratio cannot be computed for a quantitative variable from a qualitative variable (pp. 102–104); that probability statements about sample statistics can also be properly interpreted as probability statements about population parameters (p. 143); that probable errors should be used for standard deviations and correlation coefficients (pp. 145–146); that the t-test should be used to test whether or not a sample could have been drawn at random from a known population (where the population variance can be known as well as the population mean) (p. 157); and that in most instances the number of degrees of freedom for t is just one less than the size of sample (p. 159). The biserial and tetrachoric correlation coefficients are introduced as new measures of correlation rather than as equivalent to the product-moment coefficient; their underlying assumptions are not stated; and indeed the examples used do not fulfill the assumptions (pp. 108–110). The connection between partial correlation and prediction is misconstrued by introducing a vague notion of a "real" correlation instead of using the definition of correlation used earlier (pp. 127–128). In general, the assumptions behind the various significance tests are not given, especially those of analysis of variance and covariance, although—in contrast to other psychologists—the author introduces the normal distribution where it belongs, namely in connection with sampling problems.

Deficiencies noted in the discussion of graphic presentation include the following: Grid lines are drawn through the bars of a histogram (p. 9); in another there is a mixing of light and heavy lines for bar outlines (p. 16). The impression is given that the median is graphically represented by a line segment, whereas the mean is shown as a point (p. 16). Pie charts are endorsed liberally for complicated comparisons (pp. 43–44). Vertical "frequency" scales are retained for a frequency polygon and a smoothed curve, instead of using an area key (pp. 63–64).

REVIEW BY HELEN M. WALKER
Professor of Education
Teachers College, Columbia University

THE attempt to explain the meaning of commonly encountered statistical terms without using any formulas whatever and using almost no compu-

tations is a task requiring considerable courage as well as great skill in exposition. This book has been more nearly successful than the reviewer anticipated it could be, knowing only too well the difficulties to be encountered. The language is clear and simple; the illustrations are well chosen and within the understanding of the sort of person likely to read the book; almost all of the terms ordinarily used in non-theoretical studies are included although the coverage is more complete for education than for biology, business, or economics; no previous preparation is presumed except an alert and inquiring mind.

The first major problem to be faced by the writer of such material is the choice of topics. Shall he treat the topics the reader is likely to come upon in published research or only those topics which he considers appropriate to use? If he omits certain outmoded statistics, the person who reads his book in the hope of finding enlightenment on the meaning of published research is doomed to disappointment. Dr. Butsch apparently decided to include in the first part of his book (and to treat with respect) statistics which are far less efficient and more ambiguous than other more modern statistics which he treats in later chapters. The reviewer would question this decision.

The topics referred to in this text include: frequency distribution (6 pp.), central tendency (8 pp.), graphic representation (38 pp.), all of which can be presented easily on a non-technical level and have been so treated in a number of introductory texts; measures of dispersion (10 pp.), including quartile deviation, mean deviation, standard deviation, with two computations of each from easy data; the transformation of raw scores into percentile scores and standard scores (12 pp.); simple correlation and regression (32 pp.), including brief mention of the correlation ratio, bi-serial correlation, the coefficient of contingency, and tetrachoric r; multiple regression and multiple and partial correlation (21 pp.); a chapter called "Measures of Reliability" (15 pp.), which brings in such terms as sample and population, statistic and parameter, random sample, sampling error, sampling distribution, standard error, "determining the limits within which a parameter will probably be found," probable error; a chapter on "Measures of Significance" (17 pp.), which brings in such terms as the probable error of a difference between means, critical ratio (i.e. a difference divided by its probable error), the t-test, degrees of freedom (on which a fair amount of light is shed in less than a page), the null hypothesis, and levels of significance; and a second chapter on measures of significance, including the chi-square test (5 pp.), treated without reference to the previous mention of the coefficient of contingency and with an example using a 2×2 table of only 40 cases in which Yates' correction would have made considerable difference had it been applied, analysis of variance (7 pp.), and analysis of covariance (6 pp.).

After the topics to be treated have been determined upon, it is indeed a difficult task to make them take on meaning without either computational exercises or mathematical derivations. In fact, it is practically impossible to be intelligible on an elementary level and also completely correct if one is

dealing with topics beyond the simplest ones. If the writer of such a text is primarily concerned with making statements which cannot be misconstrued and which cannot be called in question by the person who knows his theory, he will hedge his explanations with so many conditions and qualifications that the reader is left in a haze. Therefore it seems pedantic to point out incorrect statements and inconsistencies in a book which on the whole creates a very favorable impression, but nevertheless it seems necessary to mention a few such.

There are about 10 pages devoted to the probable error written in the manner of 20 years ago, even including such a statement as the following: "Thus, if a correlation coefficient for a particular sample is reported as .60 ±.02 we know that the chances are even that the correlation for the total population will not be less than .58 or more than .62; and the chances are very remote that it might be as small as .52 or as large as .68." After reading this, it is a real surprise to come upon a rather good elementary treatment of *t*. The spirit and even the phraseology of these two sections are so dissimilar that it is hard to understand how the author who wrote one of them could also have written the other. The relation between the critical ratio and the *t*-test is left in some confusion. The reviewer feels that the extended treatment of probable error with all the outmoded phraseology of interpretation is unfortunate and that it would have been better to make the whole explanation in terms of standard error with a sentence or two explaining the relation between standard and probable error and a statement that the latter could have meaning only for a statistic that was normally distributed. The implication of this section and the suggestion on page 147 that the probable error can be applied to a difference between percentages, standard deviations, or coefficients of correlation seems to suggest that all sampling distributions are normal although that incorrect statement is never made. The term, "sigma," is used throughout instead of the more straightforward "standard deviation." Why authors do this, substituting the name of a Greek letter (which is no longer widely used as the symbol for the standard deviation of a sample) for a perfectly good noun, the reviewer never can understand. The treatment of partial correlation is particularly happy though it might have been made still better by the use of the concept of residual error. The illustration of the *t*-test on page 157 seems to suffer from errors in printing and will be most confusing; it refers to ten scores but prints only nine. However, the mean as stated is not the mean of this nine. Even by adding a tenth score, which would produce the mean quoted, it is impossible to secure the value of *t* quoted, and the reviewer cannot decide what computations were actually made. Certain topics such as the correlation ratio, bi-serial *r*, and tetrachoric correlation are treated so briefly that one wonders if their introduction will not be confusing. The book might have been made more useful by the inclusion of well selected annotated references to more extended treatments available elsewhere.

The person making his first acquaintance with statistical method will gain

much by a cursory reading of this book. It can be recommended to one who wants only to know the general spirit of statistical investigation, and that is the person for whom the book was written. If subsequently this person undertakes to use the intentionally brief discussions of this book as a guide for writing an interpretation of data, if he tries to press the meaning of individual sentences too far, he will be trying to make the book serve a purpose which it cannot fulfill and which it was never intended to fulfill. The writer has designed the book for "a large number of teachers, social workers, personnel directors, and industrial executives, who through lack of opportunity or inclination have failed to acquire a minimum of statistical techniques"; many of whom, however, "should be able to read the articles employing statistical procedures and should at least be able to understand the purpose and the results of the techniques employed." For such persons this book should be useful.

Rudimentary Mathematics for Economists and Statisticians. *W. L. Crum* (Professor of Economics) and *Joseph A. Schumpeter* (Professor of Economics). (Harvard University.) A revision of Supplement to *Quarterly Journal of Economics*, Vol. 52, May 1938. New York 18: McGraw-Hill Book Co., Inc. (330 West 42nd St.), 1946. Pp. xi, 183. $2.50. *Two reviews follow:*

REVIEW BY J. E. MORTON
Cornell University

PERHAPS the most outstanding feature of the book is the pedagogical skill with which the student is introduced to and led through the field of calculus. The ease of exposition, and the smooth and well-integrated development of the idea of rates of change and their systematic treatment, result in a volume which places all teachers considerably in the debt of the authors. From this point of view the text compares favorably with such simple and successful introductions into the pattern of thought of the calculus as DeMorgan's and Irving Fisher's.

The present volume covers very much the same ground as the well-known original, which was published as a supplement to the *Quarterly Journal of Economics* in March 1938. The new material includes paragraphs on Taylor's series, on homogeneous production functions, on Lagrange multipliers, on the concept of the integral, and an enlargement of the treatment of higher order differential equations. A brief discussion of determinants is added in a new chapter. The book should make easy and pleasant reading for any student who still remembers the essence of his high school algebra course.

The less satisfactory aspects of the book are the absence of bibliographical material and the absence of problems and examples, which some instructors or readers may deplore. The book addresses itself primarily to the future economist rather than to the statistician.

There is no doubt in the mind of the present reviewer that the book will be sincerely welcomed not only by teachers but also by those students in

economics who do not feel quite up to the task of working their way through, say, Allen's *Mathematical Analysis for Economists.* Projected against the background of present-day textbook literature, *Rudimentary Mathematics for Economists and Statisticians* represents what may be considered the prelude to the inclusion of mathematics as part of the economics curriculum. Professors Crum and Schumpeter in their foreword seem to suggest such a claim and the book provides ample justification for it.

In trying to evaluate this book, two questions come to mind: (a) Is there need for a book of the nature of Crum and Schumpeter's *Rudimentary Mathematics for Economists and Statisticians?* (b) Since it is written in the fashion of a textbook, or at any rate, in such a way that it would lend itself admirably to use as a text, should the existence of such a book lead to the establishment of a course of a similar nature, and is such an addition to the conventional economics curriculum desirable?

If the book is used as collateral reading or for purposes of self-study, the answer to the first question is an unreserved "Yes."

As to the second of the two questions, it will not be easy to reach a satisfactory answer. There can be little doubt that mathematics and mathematical models of thought are being used to an increasing extent in economic literature. This emphasis on economic analysis by quantitative methods is no longer limited to the loftier areas of the graduate curriculum, but it has been introduced into the rather pedestrian regions of the introductory economics course. At the same time, the training of economics students in the understanding and the use of mathematical language is sadly lacking. The question of how to remedy this situation is certainly one of the most important of those confronting the planners of economic curricula. The problem is a pedagogical and an administrative one, permitting many widely different answers. Should the student in economics undergo a more thorough training in mathematics, say, through the calculus as offered by the existing mathematics departments? Is such an intensive training, as an integral part of a student's education, feasible for, say, the student in science but not for the young economist because of "lack of time" or any other pretext or reason?

This latter belief, although it is rather popular, requires careful thought. As in the case of the teaching of statistics to economists, the point of view prevailing at the time of introduction of a new course into the curriculum may not prove to be altogether happy and desirable in the long run.

REVIEW BY GERHARD TINTNER
Professor of Economics and Mathematics, Iowa State College

THIS book provides an elementary introduction to some methods of mathematical analysis for economists. The first and second chapters deal with graphic analysis and related subjects in analytical geometry and simple algebra. They are introduced with the help of a discussion of some problems in cost theory. Chapter 3 treats the concept of the limit. The examples are

taken from the theory of marginal cost, of demand curves, of marginal utility and of marginal revenue. Chapter 4 discusses rates and derivatives. This includes various rules for differentiation, which are, however, introduced without proof. Higher derivatives, Taylor series, partial and total derivatives are also included.

Chapter 5 takes up maxima and minima in one and in several variables as well as a brief treatment of inflection points. Examples are given from cost and production theory as well as from the statistical theory of linear regression. Chapter 6 deals with differential equations and their integration as well as with definite integrals; also discussed are elasticity of demand, compound interest, total utility and some simple problems in economic dynamics. The last chapter gives an introduction to determinants.

The authors ought to be congratulated for having achieved almost a miracle of condensation. A great number of difficult subjects are discussed in a very small space. This could only be achieved by presenting many mathematical theorems without proof.

This lack of proof may, however, be confusing to the reader. It would have been possible to give, for instance, exact proof of the rule for the differentiation of a power if the binomial theorem had been introduced. The authors should have included this theorem also because of its great importance for the student of statistics and probability. On the whole, the needs of the student of statistics are somewhat neglected; the only statistical subject used as an example comes from the theory of simple regression. Other proofs for rules of differentiation could also have been given without too much trouble and without adding excessively to the bulk of the book.

This book certainly fulfills a very useful purpose as an introduction to calculus for the somewhat mathematically inclined and interested economist. It is particularly recommended because it is free from the preoccupation with physics and especially classical mechanics which characterizes almost all introductory calculus books. A knowledge of these subjects is of no great importance to the economist. He will save himself time and unnecessary trouble if he uses this introduction to calculus which utilizes exclusively examples from economic theory and mathematical statistics.

If we compare the present book with other available introductions to mathematics for economists, it appears that it is somewhat inferior to R. G. D. Allen's *Mathematical Analysis for Economists* which is generally considered the classical treatment of the subject. Allen's book covers a much wider range of mathematical theory and also gives an almost complete survey of the traditional theory of mathematical economics. Another book, *Elements of Mathematics for Students of Economics and Statistics*, by D. C. Jones and G. W. Daniels (University Press of Liverpool, London, 1926) is also somewhat preferable. But both of these books are written for English students who are able to approach college economics with a better high school education in mathematics. Hence, it seems that they are not as suitable for teaching purposes as the present book.

All the examples included are artificially constructed. One may wonder why the authors have not used some of the results of recent empirical econometric research. These would provide excellent examples for an introduction to mathematical and statistical methods. The authors could have used, for instance, some of the empirical demand curves established by Henry Schultz, the production functions of Paul Douglas and his school, and the cost functions established by Joel Dean—to mention only a few possibilities. These empirical results should take the place of the artificially constructed examples. They could be expected to appeal more to the students of economics and statistics and stimulate in them an inclination towards empirical research.

On the Solution of Normal Equations and Related Topics. *David B. Duncan* (*Pawlett Scholar, University of Sydney*) and *John F. Kenney* (University of Wisconsin at Milwaukee). Milwaukee 3, Wis.: Book Store, University Extension Division (623 West State St.), 1946. Pp. iii, 35. $1.00. Paper.

REVIEW BY WALTER L. DEEMER, JR.
Lt. Colonel, Air Corps; Chief, Department of Statistics
AAF School of Aviation Medicine, Randolph Field, Texas

THE preface to this monograph states that it is "intended to serve as an introduction to modern computational methods. . . . We explain the matric theory that is useful in this connection and then describe some of the 'compact' methods which have been developed recently, including the useful 'square root' method."

Actually, it is an introduction to modern methods of computing the inverse of a matrix and to computing the solution of a set of simultaneous equations. It is not an introduction to modern computational methods in general.

What the monograph does is to give in 35 pages a very clear and detailed explanation on an extremely elementary level of how to solve a set of simultaneous equations and how to get the inverse of a matrix; it also gives the matrix theory, so that with a very small amount of study a beginner can learn to do the computation quickly and accurately and can also learn why each step is taken.

For someone with no knowledge of matrix algebra, this monograph is an excellent introduction to the general concept of matrix theory and its application to the solution of simultaneous equations. A thorough study of this book will give even the beginner enough background to proceed to more advanced material on matrix algebra.

The first two pages explain how simultaneous equations arise in least squares theory. This is followed by twelve pages on the more elementary properties of matrices and determinants. The general concept of factoring a matrix as a preliminary step to finding the solution of the equation or to finding an inverse is then developed. The next fifteen pages are devoted to an explanation of the square root method of solving equations and getting an inverse. This part of the monograph is particularly good. The exposition

is very complete, and enough detail is given, with numerical examples, so that a beginner can follow it with little trouble. This is followed by two very useful pages on the method of moving decimal points to facilitate computation. The pamphlet ends with two pages on nonsymmetric equations.

The authors develop no new theory and make no claim to originality. They have quite wisely refrained from trying to explain the wide variety of methods now available for computing an inverse. They have chosen a method which they believe to be one of the best, and have explained that method in detail. This reviewer believes that for routine calculation by clerks there are better methods than the square root method, but even if this is true, it is not a serious flaw in this monograph since after studying this book the beginner may go to the literature (there is a fairly complete bibliography given) and make his own evaluation of methods.

It is hoped that a consolidation of the periodical literature on the theory and practice of solving simultaneous equations will soon appear. When it does, this book by Duncan and Kenney will be an ideal companion to it.

Unfortunately, there are a number of misprints. Some of them will cause no difficulty even to the beginner. Others may cause the beginner some trouble, particularly those which occur in numerical examples where cross reference between the data and the operations of solution is made difficult by gross errors in transcription. A number of such errors occur on pages 18 and 20. A list of misprints found by the reviewer follows: p. 6, line 4, for *column* read *row;* p. 6, equation 3a, for G read G'; p. 13, exercise 1, for superscript *5* on MacLane read *2*; p. 14, equation 21, first equation, for g_{11} read b_{11}; p. 18, equation 20, second equation, read $(g_1-a_{1i}k_1)/a_{11}$ for $g_1-a_{1i}k_1/a_{11}$; p. 18, line 16 $(S_{21}=\cdots)$, read .183 for .777; p. 18, line 19 $(k_1=\cdots)$, read 2.03 for .777, add to end of expression $=.777$; p. 19, line 10 from bottom, for (0.023) read (−0.023); p. 20, line 17, for A read G; p. 20, line 18, for A read G; p. 20, line 6 from bottom, for .777 read 2.03; p. 20, line 5 from bottom, for −.037 read +1.00; p. 25, line 8, for .132 read −.132; p. 25, line 6 from bottom, for $(.354)^2$ read $(-.280)^2$; p. 25, line 3 from bottom, for (.030) read $(.039)^2$; p. 30, exercise 1, the matrix given for A^{-1} should have the decimal point of each element moved four places to the left; the values of the u_i are correct as given; p. 32, line 12, for x' read x; for U' read U; p. 32, line 14, for 10 read 10^{-1}; p. 33, line 11 from bottom, for P read R.

Government Statistics for Business Use. Edited by *Philip M. Hauser* (Assistant Director, Bureau of the Census) and *William R. Leonard* (Deputy Chief, Division of Statistical Standards, Bureau of the Budget). (Washington, D. C.) New York 16: John Wiley & Sons, Inc. (440 Fourth Ave.), 1946. Pp. xvi, 432. $5.00.

REVIEW BY WILLIAM A. SPURR
Professor of Business Statistics, Stanford University

THIS book tells what information is available from the Federal Government, the agencies from which it can be obtained, and ways in which

it can be applied to business and economic problems," according to the jacket. More precisely, it is a descriptive reference book for business men, somewhat incomplete and loosely organized, but nevertheless a valuable guide to the principal types of government data useful in management, production and marketing. It is written by twenty government statisticians (seventeen of them from the bureaus of the budget, census, and foreign and domestic commerce) who are leaders in their fields. The book is properly divided by subject matter rather than by government agency. Chapter headings include general business indicators, manufacturing, mining, agriculture, wholesale and retail trade, international trade, public utilities, corporate financial statements, money and banking, prices, population, housing, and labor.

The treatment is primarily descriptive, with few tables or charts of the data themselves. A useful distinction is made in some chapters between "benchmark" statistics (chiefly census figures), current data, and special surveys. War agency statistics of ephemeral value are properly omitted. Units are carefully defined in many instances (for example, "wholesale prices," page 308). Some of the writers also describe the gaps and defects in the available figures.

A praiseworthy effort is made throughout to show the general uses of the data in business, chiefly in general management and marketing. Sometimes specific applications are also cited. These range from actual cases, such as that of a publisher who matched subscription lists against census schedules (p. 357), to more unrealistic suggestions, such as that the operator of a small grocery store should keep a ten-year monthly chart of five general price indexes (p. 318).

While this book is of average length and contains much excellent material, it is still far from complete as a source book for business statistics. Little reference is made to important private sources of data (except for minerals) on the insufficient ground that "these sources are generally well known" (p. 10). Even within government sources, "it was necessary to leave out many statistical series of direct interest to business" (p. 10). Furthermore, while some topics, such as national income estimates and the population census, are discussed in detail, many others receive such fragmentary treatment as interest rates (pp. 287–280) or "Foreign Commerce Weekly contains frequent articles on various aspects of foreign trade" (p. 109, quoted in full). The business economist or statistician may also feel that the book is "written down" for the small business man. For example, both foreword and introduction attempt to justify the use of statistics themselves, and chapter introductions point out the importance of such topics as manufacturing, trade, and prices.

A few writers understandably stress the virtues of their own bureau's data (for example, national income as a business indicator) without enough attention to the limitations of their estimates or the contributions of other agencies. Those who discuss the work of others are apt to be most objective

and critical. The sections on manufacturing, mining, financial statements, and banking are among the best in this respect.

A final criticism reflects the difficulty of editing the work of many independent contributors. There is some unevenness of treatment, such as in the inclusion of illustrative tables, war agency data, subject vs. source classification within chapters, case citations, and overlapping of subject matter. In some of the latter cases, such as retail sales (pp. 48, 163) and farm prices (pp. 114, 305), minor discrepancies indicate that the duplicating sections have not been cross-checked. The index, which is an essential aid in using a book of this kind, should be more detailed and cross-classified. The bibliography, too, might well be more complete, and classified by subject rather than source, as the book is.

Despite these limitations, the book provides both the business man and the statistician with a valuable synthesis of essential government data. With the war stimulus to fact finding (tempered by the Federal Reports Act of 1942), the current ambitious census program, and improvements in sampling techniques, the government's statistical aids to business should continue to increase. It is hoped that this volume will be supplemented in the future by a still more comprehensive and critical survey of government and private contributions to business statistics.

Changes in Income Distribution During the Great Depression. *Horst Mendershausen* (Professor of Economics, Bennington College). Conference on Research in Income and Wealth, Studies in Income and Wealth, Vol. 7. New York 23: National Bureau of Economic Research, Inc. (1819 Broadway), 1946. Pp. xviii, 108. $2.50. *Two reviews follow:*

REVIEW BY MORRIS H. HANSEN
Statistical Assistant to the Director
Bureau of Census, Washington, D. C.

THIS study of the changes in income distribution during the depression years is based primarily on an analysis of income data reported for both 1929 and 1933 for identical families. These data, which show the joint distribution of 1929 and 1933 income for a sample of families in each of 33 cities, were obtained as a part of the Financial Survey of Urban Housing taken during 1934. Although there is little information showing the incomes of identical families at different time periods, an effort is made to bring what is available to bear on the problem in order to verify the inferences drawn. The author obtains summary statistical descriptions of the data and formulates and tests hypotheses concerning the nature of income shifts during a period of economic fluctuations, with primary attention centered on a period of shift from prosperity to depression. The analysis is an intensive one of the available data.

The rest of my remarks will deal primarily with the methods employed rather than the economic significance of the findings.

It is difficult in reading the book to determine whether it is directed primarily at one interested in the methodology, or at one interested in the eco-

nomic theory involved. For one concerned with the methodology, the book should be of particular interest as it contains a full exposition of the techniques of analysis employed, and the reasons for choice of various methods. In addition, alternative methods of analysis are carried through to investigate whether a different approach to a measurement being made or an hypothesis being tested would have led to different conclusions. For example, changes in income inequality are studied both through using the coefficient of variation as a measure of inequality and the coefficient of concentration (based on the mean of the differences between individuals, disregarding the direction of the differences). In chapter 2 a particularly interesting method of analyzing changes in the sections of the income distribution is introduced.

For one primarily interested in the particular hypotheses explored and the economic significance of the analysis, many useful results are presented; but the discussions of methods and of economic interpretations are so intermixed throughout that it is harder to read than would have been the case if the methodological aspects had been concentrated more heavily in the appendices and less in the text.

The author was limited to data already available to him from existing sources. As a consequence, he had to use such data with whatever defects it might have for his analytical purposes. The sampling carried out in 33 cities in the *Financial Survey of Urban Housing* was subject to a number of biases. One can never be completely sure in analyzing such data exactly how valid his conclusions will be, but the author attempts to determine and indicate the probable nature of any biases in his results caused by the kind of original data he was working with, and he succeeds in making effective use of the available information.

While high standards of analysis are followed for the most part, there is at least one major point in which a serious misinterpretation has been made, and a number of minor points in which the analysis may be questioned. The principal problem is in chapter 1 in the analysis of changes in the average income level of cities. The author draws the conclusion based on a regression analysis that cities with low income level in 1929 have a greater relative decline in income than cities with high income level in 1929 (see pages 19 and 20). He shows that the regression of 1933 average income, Y, on 1929 average income, X, has nearly the same slope as the line through the origin having a slope of the ratio of the means, but the regression of X on Y has a very much steeper slope with respect to the X-axis. He properly concludes that these two lines would have been identical had there been a perfect correlation between X and Y. However, he concludes that lack of a perfect correlation arises from errors in the observations and that the true line would lie somewhere between these two fitted lines and that cities with low income levels in 1929 had a greater relative decline than cities with high income levels. I think that the analysis and the conclusion in this case are faulty. If the only reason that individual cities did not fall precisely on a straight regression line was because of errors in measuring the average

incomes in those cities, then the analysis made would be correct. Actually, however, if the average city incomes were known precisely for each city for each year, there still would be a dispersion about the line and probably almost as much dispersion as is observed in his data. The errors in the observations probably contribute relatively little to the lack of correlation between 1929 and 1933 average incomes. If this is true, the appropriate regression to have used to compare the average income level in 1933 of cities having a given average income in 1929 is the regression of 1933 average income on 1929 average income.

A few other relatively minor criticisms should be made of his remarks on methodology.

On page 23, in discussing the measurement of income inequality, the relative merits of the standard deviation and of the mean of all possible differences between individual families are compared as measures of inequality. It is remarked that the standard deviation is of special utility because of its high sampling stability. This high sampling stability, in fact, is an advantage of the standard deviation in sampling from normal and certain other populations, but the standard deviation may or may not have higher sampling stability than other measures of inequality when sampling from abnormal distributions such as the income distribution.

In a few instances throughout the volume, tests of significance are made separately for each of the 33 cities, and the conclusion is stated in terms such as, on page 79, "the correlation is statistically significant on the 5 percent probability level in the three cases that show the larger coefficients of correlation." The individual tests of significance based on the larger differences observed among a set of observations are improper.

Again, on page 104 and the top of page 105, the statement is made that a proper test of discrepancies of certain observations from a line of regression would entail an analysis of variance for each city sample but that for technical reasons this type of analysis was not carried out. The technical reasons cited were that the distributions were far from normal. Actually, the type of test proposed will be a valid test when dealing with large samples whether or not the individual observations are normally distributed, and large samples were being used in this instance. However, the process that was followed for dealing with this problem was quite adequate so that no confusion in the analysis or in the conclusions that were drawn resulted.

There are other minor errors that have no significant effect on the conclusions from the analysis. One or two slips in the text may cause confusion. On page 100, formula (1; 2) should be σ^2 instead of σ. On page 103, it is stated that "It can easily be shown that the area of the polygon *AEFG* is: $\frac{1}{2}\Sigma r_i(q_i+q_{i-1})$"; actually the polygon referred to should be *ADGFE*.

On the whole, the analysis is a careful and scholarly one and contains an excellent beginning in a field where more analysis and improved analytical tools are needed.

REVIEW BY JOHN LINTNER
Assistant Professor of Finance, Harvard University

THIS book is a valuable addition to the series of studies of income and wealth by members of the staff and consultants of the National Bureau of Economic Research. In the present volume, Dr. Mendershausen adds very substantially to our knowledge of fluctuations in the size distributions of family money incomes between prosperity and depression. The study is based primarily on an analysis of the valuable data on incomes of identical families in 1929 and 1933 supplied for 33 large and middle-sized cities by the Financial Survey of Urban Housing. Information from this source is supplemented by income tax statistics for the United States, Wisconsin and Delaware, and earnings data from the Social Security Board and German income and wage data. Dr. Mendershausen's work is thus based on a broader range of data, better suited to his purpose of analyzing the changing structure of the entire range of the income distribution, than has been used in any previous study. The material is handled in a careful and competent manner. Moreover, the whole analysis is enriched by a considerable insight into the economic forces which produce the observed changes in the over-all distribution of incomes, and the interplay of changes in different parts of the income structure. The important new information and analytical results presented in this study, as well as the suggestive discussion of their significance, and the problems raised for further investigation make this work important reading for all economists. While at many points the exposition is involved and difficult, readers of the study will be well repaid for their effort.

Economists are concerned with changes in the distribution of income receipts because of their significant welfare implications, and also because shifts in the distribution of income may be expected to affect the volume of aggregate demands for consumer goods as well as the volume and composition of new saving. Through these factors, changes in the *distribution* of income will influence the *level* of total income payments, business activity, and employment. Shifts in the distribution of income influence aggregate demands for consumers' goods and the volume of saving because different income groups tend on the average to spend differing proportions of their income on various goods and correspondingly to save differing proportions of their income. In order to refine their estimates of the total demand for consumers' goods and the level of savings that will tend to be associated with different levels of total income payments, economists therefore need to know: first, how the total income will be distributed between different income groups; and second, how these different income groups will tend on the average to allocate their income between different types of saving and expenditures on various consumers' goods and services. Mendershausen's work is concerned exclusively with the first of these two types of information.

His important contributions in this respect are three in number. In the

first place, his results indicate that, contrary to previous impression, the relative dispersion or "inequality" of the income distribution as a whole tends to vary *inversely* with the level of total income payments over the cycle. Incomes tend to be relatively more unequally distributed in depression than in prosperity. The Financial Survey data, dealing with the decline from 1929 to 1933, show that in most all of the cities income inequality was greater in the latter year, and the supplementary data studied tend generally to show a similar tendency in periods of falling incomes. These supplementary data likewise generally show a decline in the inequality of the entire distribution in years of rising income, although in this case the exceptions are somewhat more frequent. Some additional support for the hypothesis that income inequality declines with rising incomes may be found in the recent sampling survey conducted by the Department of Agriculture and the Federal Reserve Board, which indicates that for the United States as a whole the Gini index of inequality stood at .38 in 1945 as compared with .47 in 1941 and .46 in 1935–36. However, the apparent failure of inequality to decline between the latter two years again casts some doubt on the validity of the hypothesis.

In this connection, Mendershausen goes on to show that the general tendency he finds for the inequality of the entire income distribution to vary inversely with the level of total incomes is perfectly consistent with the findings of earlier studies that (*a*) the inequality in the distribution of incomes over \$5,000, and (*b*) the share of the top 1%, 5% of income recipients in total income payments both vary *directly* with the level of aggregate incomes. Indeed, Financial Survey data show both these latter characteristics which relate, however, only to changes in the inequality of the upper part of the income scale. As Bortkiewicz had shown, changes in the inequality of the entire income distribution depend on the (weighted) changes in inequality within *both* the upper *and* lower groups of incomes as well as the changes in the relative mean incomes between the groups. Without exception, Mendershausen's data show that the inequality within the lower 50%–70% of income recipients was greater in 1933 than in 1929, and that the relative difference in mean incomes of the two groups had increased. These two effects were sufficiently strong in most cases to outweigh the reduced inequality among the upper range of incomes. Inequality of the whole distributions consequently increased. This analysis is important because it emphasizes the danger of imputing changes in part of the distribution to the whole range of incomes, and because Mendershausen's generalization, if later confirmed more fully, would bring the empirical results of Polak's and other studies of the effects of changing inequality of incomes on consumption expenditure into agreement with the established body of economic analysis.

Mendershausen's second major contribution lies in his effort to explain the observed changes in the quantitative characteristics of different segments of the size distribution of total family incomes by establishing patterns of

change within the upper and lower ranges of the separate distributions of different qualitative types of income. For instance, the greater inequality among lower incomes during depression is traced back to (*a*) the expansion of unemployment, (*b*) the greater importance of the income gap between the employed and the unemployed, (*c*) the unequal incidence of unemployment among the lower and the more highly paid workers, and (*d*) the increasing mean difference in the wage rates of high and low-pay jobs. This analysis represents a significant advance in our understanding of changes in the lower part of the income distribution.

Similarly, Mendershausen explains the changing distribution of the upper groups of income in terms of (*a*) the tendency for income from property to fluctuate cyclically more than income from work and (*b*) the fact that very high incomes usually include more income from property than income from work. While this explanation is probably acceptable in very general terms, it is marred by the fact that property incomes are made up of such heterogeneous types of return as dividends, rents and royalties, interest, capital gains minus capital losses, and "other income." These different types of property income are known to have different patterns of change over the cycle, and the entire explanation of the observed changes in the distribution of large incomes would have been substantially improved if these differences in the behavior of different types of property income had been studied and introduced into the analysis.

This effort to establish patterns of change in the distribution of different qualitative types of incomes represents a most constructive step. Mendershausen simply doesn't push his analysis of the behavior of different factoral incomes far enough, particularly in the case of property incomes. Nevertheless, he presents a very suggestive and helpful analysis as far as he goes, and it is to be hoped that his work will stimulate further study along these lines. Surely this type of analysis is an indispensable supplement to the study of "income generating functions" suggested by Marschak in his preface.

The third important contribution of Mendershausen's study is to provide for the first time in the literature a study in quantitative terms of the changes in the relative position of individual families in the income scale between prosperity and depression. Everyone's income does not change in the same proportion over the cycle: some families in each income group in a base year gain in relative position, and others drop behind. Data on these shifts are provided by the Financial Survey in the form of a *joint* distribution of sampled families by their income in both 1929 and 1933. Mendershausen finds that both the absolute and relative dispersion of the 1933 incomes of those families in the highest and lowest income groups in 1929 are greater than for the 1933 incomes of the central income groups of 1929. Interestingly enough, the diversity of the high-income families of 1933 with respect to their former incomes is less than the dispersion of the top-income families of 1929 with respect to their subsequent incomes. Perhaps the most significant findings relate to the analysis of "favored" and "disfavored" groups

—those income groups whose mean income in the second year is greater or less than would have been expected if allowance were made only for changes in the level of total incomes, the changed inequality of the total distribution, and the shifts in rank of the distribution as a whole. On the basis of this standard, the recipients of very low and moderately high incomes in 1929, fare relatively better during the depression than those with moderately low and extremely high incomes in 1929. These differences are again explained in terms of the relative stability of different qualitative types of income return.

This analysis of the changing relative position in the income scale breaks new and significant ground; but more studies will be needed to test the stability of the pattern shown for the 1929-33 experience. Moreover, the full value of this type information cannot be realized until budget data are available for the average expenditure patterns of members of given year income groupings cross-classified by previous years' incomes.

Five Week Patterns of Prices and Volume: A Classification of the Weekly Movements of Trading Velocity and the Dow Jones Industrial Averages through the Twenty Year Period 1926-1945. *Arthur A. Merrill.* Schenectady 8, N. Y.: the Author (1567 Kingston Ave.), 1946. Pp. ii, 32. $1.00. Paper.

REVIEWED BY OWEN ELY
Editor, The Analysts Journal
61 Broadway, New York 6

This 34-page pamphlet consists of a two-page explanatory preface and 32 pages of tables. The author has attempted to determine whether it is possible to forecast the weekly trend of the stock market by studying (1) the up or down changes in the Dow Jones Industrial price average, and (2) the changes in trading velocity (as measured by shares per trading hour), in each of the five preceding weeks. Only changes of trend were considered, without regard to the amount of change. The study covers the period January 1, 1926 to May 30, 1946, which includes 1,063 five-week periods (one beginning each week). The varying pattern of changes in price and volume for each five-week period have been studied in relation to the trend of prices in the week immediately following.

There are 32 basic patterns, representing the number of combinations of up and down for a five-week period. Each page is thus devoted to a single price pattern, which appears at the top of the page. Each price pattern may be combined with 32 potential volume patterns of the same character, and all of the latter are repeated on each page. For example, page 1, item 1, pictures a five-week period in which both price and volume gained steadily; while on page 32, item 32, both prices and volume were lower in each consecutive week. In between there are 1,022 other combinations of price-volume patterns. Opposite each pattern there is indicated the number of times the market advanced in the following week and the number of times it declined.

Evidently this system proved somewhat disappointing, for Mr. Merrill decided to add the results when volume patterns varied (from the one pictured) with respect to one week out of the five. Thus the number of potential forecasts would be nearly doubled. In processing the results for forecasting, it was obviously necessary to select only outstanding results where price changes (in the week following the five-weeks test period) showed a preponderance of ups or downs. As to the value of the forecasting results, the report seems rather diffident and incomplete. The writer in the last year and a half found that he had "good luck" where the difference in the number of ups and downs was two or more, or where there was a difference of five or more for the "near" volume patterns. In this period there were eighteen correct forecasts and four incorrect, so that the forecasts obtained were 83 per cent accurate; but since 17 per cent would involve losses, the net favorable result would be about 67 per cent. However, the failure to report any data on the actual amounts of gains and losses, with allowance for dividend income and cost of trading (including taxes), makes the conclusion mainly of theoretical interest.

Moreover, applying the same rules (differences of two and five or more, in the actual and near results) to the entire period covered by the book, 1,063 weeks, definite forecasts were apparently obtainable in only 191 weeks or 18 per cent of the time, as compared with 30 per cent in the short period reported by the author. No attempt was made to work out the value of the forecasts over the whole period as this would involve considerable study by this reviewer.

This system of forecasting is purely mechanical; it appears likely that much more valuable results would be obtainable if the *amount* of change in price and volume, instead of merely the direction of the change, could be used in a forecasting formula. This would, of course, involve a tremendous number of computations; but the problem would be relatively simple for some of the wonderful new computing machines now available. Even the use of IBM equipment, with punched cards, would solve the mechanical problem (Mr. Merrill's tabulations were prepared "with the aid of a home-made calculating machine").

The number of research projects which could be developed to aid forecasting stock market trends is limited only by the number of permutations and combinations of published data relating to the stock market—or by the energy and time possessed by the students of forecasting techniques. The results stated in Mr. Merrill's study are so fragmentary and incomplete that his little brochure has little practical interest, except that it may open up a new field of study for students of the market.

LETTERS ABOUT BOOKS

Readers are invited to submit letters about statistical methodology books for publication in this forum. Concise, informative letters which supplement previously published reviews by pointing out specific strengths, weaknesses, errors, and errata in currently used books are wanted. Criticisms based on actual use of a book as a text are especially desired from statistics instructors. Other letters may consist of suggestions for the writing of books and reviews. Letters which contain adverse criticisms of Journal reviews will be submitted to the author of the review for any reply he may care to make. Contributors are requested to avoid personalities. The right to decide whether a letter merits publication is reserved. Letters should be sent to the review editor, Oscar K. Buros, Rutgers University, New Brunswick, N. J.

FREQUENCY ARRAYS

More than a decade has passed since I first came upon a copy of the 48-page booklet by H. E. Soper entitled *Frequency Arrays: Illustrating the Use of Logical Symbols in the Study of Statistical and Other Distributions*, which was published by the Cambridge University Press in 1922. What Soper calls "arrays" are more commonly known today as "generating functions," a terminology introduced by Laplace, who employed them systematically in his great work, *Théorie Analytique Des Probabilités* (1st edition, Paris, 1812). So far as I have been able to determine, *Frequency Arrays* was the first elementary exposition of generating functions written in English expressly for students of mathematical statistics. At any rate, it was the first exposition of this subject that I came upon, and I was eager to own a copy. I soon found, however, that it was generally regarded by booksellers as being out-of-print, and that virtually none of my statistical friends in the United States knew of this booklet.

During my visit to England, 1935–37, I continued my inquiries about *Frequency Arrays*, and learned ultimately that possibly a copy might be obtained from L. Reeve & Co. Ltd., Publishers, in Ashford, Kent. This proved to be the case, and prior to World War II, three or four of my statistical friends and I obtained copies from this source. In May of this year I wrote to the above publishers to inquire regarding the possibility of obtaining additional copies of *Frequency Arrays*, since my statistical friends continued to exhibit a desire to purchase copies of this booklet. In reply to my letter, the publishers stated that "as we have a fairly large stock of this work still left we can supply further copies if required."

Frequency Arrays was reviewed in this Journal in 1923 (Vol. 18, pp. 1073–1074) by John Rice Miner, and contains a résumé of Soper's symbolic method of treating probability-generating and moment-generating functions. The review concludes with the following paragraph:

> "This résumé will suffice to give the reader some idea of the methods involved. Mr. Soper then illustrates their use by applications to the binomial, Poisson, Gaussian, exponential, gamma type and hypergeometric frequencies, as well as to geometrical distributions and problems of random migration. In this, as in other fields, the symbolic method proves itself a powerful engine, but whether it will come into more general use is for the future to determine."

In spite of this favorable review, *Frequency Arrays* appears to have been fairly widely ignored in the United States. On the other hand, in Soper's obituary, by M. Greenwood, published in the *Journal of the Royal Statistical Society* (Vol. 94, 1931, pp. 135–141), we find the following:

> ". . . the work which will surely keep his memory green when those who knew and loved him are forgotten is *Frequency Arrays*. The Royal Statistical Society has at least the credit of proving an exception to the rule that the prophet is not without honour save in his own country. It is true that in other English journals, *Frequency Arrays* was spoken of politely. But the only notice in an English journal which I have seen conveying the impression that Soper had made a first-rate con-

tribution to statistical methodology appeared in our *Journal* over the initials L. I. (1923, LXXXVI, 97), and the Council in awarding the Guy Medal made mention of this tract. Continental statisticians did not confine themselves to vague politeness. The French Academy of Sciences awarded Soper a Montyon Prize, and the late Professor Tschuprow wrote a review (*Nordisk Statistisk Tidskrift*, 1924, III, 414-17) in which he characterised the tract not, as some other reviewers hinted, as an ingenious way of setting out what everybody knew before, but as 'einen namhaften Beitrag zur formal mathematischen Ausrüstung des Statistikers'."

Today there is another little book by a British author which, by virtue of having some exercises for the student to work, is probably better suited than *Frequency Arrays* to class-room instruction in generating functions. I refer to A. C. Aitken's *Statistical Mathematics* (New York: Interscience Publishers. 1st edition, 1939; fourth edition, 1945. $1.50), which was reviewed by S. S. Wilks in this JOURNAL (Vol. 36, 1941, pp. 148-149). Nevertheless, in my experience, teachers of mathematical statistics, and others, seem still to find *Frequency Arrays* a stimulating little volume, and many desire to own a copy.

Therefore, I take pleasure in informing readers of this JOURNAL generally that copies of *Frequency Arrays* can still be purchased from L. Reeve & Co. Ltd., Sankey House, Brook, Ashford, Kent, England; also from the Abbey Garden Press, 132 West Union Street, Pasadena, California. The English price is 3s. 6d., postage 2d.; the American price is 85 cents postpaid.

CHURCHILL EISENHART, *Associate Professor of Mathematics and Biometrician, University of Wisconsin; Principal Mathematician, National Bureau of Standards*

A NOTE ON STATISTICAL BOOKS

IN REVIEWING statistics books it is customary to discuss such topics as technical accuracy, content, organization, balance, and suitability. There is another factor which is even more basic in writing or appraising a book—the statistical creed of the author. At present there are three major interest groups in the field of statistics, each with its own "frame of reference": the mathematical statistician whose primary interest is the theory of statistics, the operating statistician who applies statistical principles and techniques to the solution of current problems of government and business, and the subject-matter specialist to whom statistics is a tool subordinate to his major interest. The prevailing pattern of statistical thinking has been, and still is, that of the latter group, although the first group is becoming more and more influential. The second group is small, disorganized, and ineffectual professionally, although it may be the nucleus of a real profession of practicing statisticians.

The conflict between the first two groups is a conflict which exists in any science between theory and practice, a conflict which may be helpful in making theory more vital and practice more efficient. The real issue in statistics, however, is between the first two groups combined and the third group, since this is a conflict between a growing theoretical and applied science, on the one hand, and an older and outmoded statistical creed, on the other. The basic issues between these two groups may be summarized as follows: The first group thinks of statistics as a theoretical and applied science in its own right. The second group does not. The first group believes applied statistics should be a full-time profession, that statistical functions should be assigned to the statistician. The second group does not. The first group sees the statistician as an expert in statistical inference and estimation based upon sampling, and in statistical inquiry generally. The second group believes that the subject-matter specialist is the only person competent to make estimates and inferences about his subject matter.

The compartmentalization of statistics into economics, sociology, psychology, education, biology, business, etc. has set up barriers to the full utilization of efficient statistical principles and practices. If the biometricians and agriculturists have done better than the others, it is simply because they were fortunate in having an R. A. Fisher working in these fields. This compartmentalism has led to a number of strange notions such as (1) Fisher's work is limited to small samples, (2) analysis of variance and the design

of inquiry are applicable only to agriculture and biology, and (3) random sampling cannot be used in the social science fields. Fisher himself has contributed to this belief in compartmentalism for the tables by Fisher and Yates are called *Statistical Tables for Biological, Agricultural, and Medical Research*, although few of the 34 tables are limited in their applications to the three fields mentioned in the title. Compartmentalism has done more. It has led to statistical confusion by introducing 57 varieties of statistical symbols, by keeping alive erroneous theory and questionable practice, and by consistently refusing to recognize that statistics is a science composed of principles rather than a tool like typing and shorthand that one can pick up in a few months. To show how slow our subject-matter specialists are in keeping up with statistics, the 1940 edition of *Encyclopedia Americana* in its lists of books under the topics of Statistical Method and Statistics carries no reference to R. A. Fisher, *Statistical Methods for Research Workers* (first edition 1925), although it contains the names of several books of decidedly lesser importance.

Compartmentalism seems to owe its existence and development to the graduate schools because their specialists discovered that statistics had a role to play in academic research. Statistics books were written in each field to meet this academic need. As time went on, it was assumed that these books were valid sources of training for *all* kinds of statistical work. The operating and research statistician in government and business knows how absurd this assumption is. The 6-credit-hour "statistician," even though officially recognized by the U. S. Civil Service Commission in 1930, is inadequately trained both in statistical theory and in statistical practice. What we need is a new approach to statistics based, not upon the patterns of academic research, but upon the needs of government and industry.

Nor is there any greater likelihood that the results will be better in the future since both statistical theory and practice are growing at a lively pace. As a matter of fact many of the problems now facing government and business and industry require an expert in statistical, especially sampling, design, not an economist or sociologist who has had a course or two in statistics. Indeed there is a very important field of professional statistical practice which seems to be neglected, for which training in mathematical statistics is a necessary, but not a sufficient, condition for qualification. In other words neither knowledge of a subject nor knowledge of mathematical statistics, by itself, is enough.

With regard to the problem of statistics books I should like to make the following suggestions:

1. The first course in statistics should deal with general principles and should cover the fields of sampling, estimation, and inference in their more elementary aspects. This course should be taken by all persons regardless of subject-matter specialization. It is analogous to the first course in college chemistry or physics; it gives the proper orientation but it is hardly more than an introduction to the subject.

2. Advanced courses in these principles, as well as their applications to special fields, should follow this course on principles. In this way a person will not become so easily the victim of rules-of-thumb which are often inefficient, but will learn to think in terms of principles applicable to problem situations. Books should be prepared along these two lines of theory and applications.

3. It is absurd to think that one can be a statistician without an understanding of mathematics at least through the calculus. Mathematics should be a prerequisite to statistical work. Prepare special nontechnical books for the laymen and others interested in the field. We should not try to meet all statistical needs in one type of book.

4. Have more books prepared by the second group—those on the firing line in statistics. For a generation we have had stereotyped textbooks prepared by academic subject-matter specialists; more recently the mathematical statisticians have moved into the market. The former group is weak on theory, the latter group is weak on practice. What we need are books that show how to apply principles to the solutions of operating problems. "Applied" statistics to date has been largely limited to academic research. Books by Fry, Shewhart, and Simon are examples of the type of book we need.

5. Write books to inform, rather than to impress, the reader. Most books are simply a hodge-podge of topics. We need an integration of statistics to give it meaning. Where mathematical derivations are given, they are sometimes so involved that the logical assumptions on which the derivations are based are hardly mentioned, let alone described. Stating assumptions in mathematical language does not seem to be enough. More attention needs to be given to statistical principles and their applicability to real problems. Then too there is a decided lack of balance in some books. For example, one book gives about 200 pages to index numbers and periodicity analysis, while another gives about 50 pages to rank order correlation, but neither devotes anything to stratified random sampling as such, a much more fundamental problem.

A. C. Rosander, *Statistical Division, Bureau of Internal Revenue, Washington 25, D. C.*

JASA REVIEW POLICY

I offer some suggestions to be taken into account in any reformulation of the preliminary statement of Journal review policy appearing in the August 1945 issue of the *ASA Bulletin*, especially the expressed objectives:

> *a)* To provide readers with a scholarly appraisal of the book for their own reading guidance.
>
> *b)* To stimulate progress toward higher professional standards among statisticians by praising good statistical writing and scholarship and by censoring poor statistical writing and scholarship.

In view of the increasing tendency to specialize, it would appear to be a significant and valid function of the Journal review to promote a catholicity of interest on the part of specialists in other branches of the science, especially through emphasis on the wider applicability of concepts and methodology characteristic of one or another specialized field. It would also seem reasonable to expect the review to give guidance to authors as well as to readers, through commentary and generalization which need not imply praise or censure of the work under consideration. Attention might be directed, not only to the individual author's errors great and small, but also to the areas still to be explored, to the syntheses still to be made, to perspectives which should become more general, to the inevitable anachronisms introduced by the mixture of new and classical viewpoints in the treatment of different phases of statistics, to the existence of important contributions in periodicals or in foreign-language literature which are usually overlooked, etc. Furthermore, there is still room for the "creative" review, which goes on to make an original contribution inspired by some aspect of the work under scrutiny.

The traditional kind of review may be of little value to the nonstatistical or semistatistical reader who has a limited technical background but is nevertheless interested in, or working in, some phase of the science. Such readers, according to the June 1946 Journal, would seem to compose the large majority of Association members—perhaps, a still larger majority if the National Roster definition of a statistician (*ASA Bulletin*, May 1945) were applied. Furthermore, many statistician members, trained before the Fisher-Neyman revolution, have an unsure grasp of the newer statistical tools and follow contemporary literature with difficulty. The elevation of professional standards depends to a large extent on the successful development of a sound but relatively nontechnical literature to meet the needs of the nonstatisticians, the semistatisticians, and the quasi-statisticians in the Association. Special attention should be given to the review of such literature, which is not specifically mentioned in the August 1945 list of four types of publications to be reviewed in the Journal.

All members, ranging from the nonstatistician to the specialist, could benefit from "review articles"—essays which summarize in systematic, integrated form the accomplishments in a branch of statistics to date, or which present more or less authoritative statements on fundamental notions of the science, general methods, etc. Such essays should be written by invitation, probably by more than one person or with the guidance of a committee of broad perspective.

Finally, it would seem desirable to construe the objectives of reader

guidance and standards elevation to include, as a constructive aspect, the formulation of "model" outlines for textbooks and, perhaps, other formal works on statistical topics. Such outlines should take account of the observations made on similar works by JOURNAL reviewers. Their preparation and periodic revision should, perhaps, be assigned to a standing committee of the Association, one which might also be concerned with standards for the training of future members of the statistical profession.

IRVING H. SIEGEL, *Chief, Economics Division, Veterans Administration, Washington, D. C.*

STATISTICAL METHODOLOGY INDEX, NO. 6

A QUARTERLY GUIDE TO CURRENT LITERATURE

Edited by

OSCAR KRISEN BUROS
Rutgers University

This bibliographic service presents statistical methodology literature (articles, books, theses, and chapters) published in 1946 to date. Anonymous references are listed first; other references follow in alphabetical sequence by first-named authors. Volume number, issue number (in parentheses), pagination, and date of issue are given for each journal reference. Stars indicate books, theses, and pamphlets; asterisks indicate publications seen by the editor of this INDEX. Authors of papers on statistical methodology are invited to send reprints to Oscar K. Buros, Rutgers University, New Brunswick, New Jersey in order to facilitate prompt and accurate listings. Foreign-language papers should be accompanied by English translations of title. Information is desired about references omitted.

Statistics at the University of Wisconsin. *Biometrics B* 2(4):70-3 Ag '46.* [653

Aiken, Howard H., and Hopper, Grace M. The Automatic Sequence Controlled Calculator. *Elec Eng* 65 (8, 9, 10, 11):384-91, 449-54, 522-8 Ag-S, O, N '46.* [654

★Aitken, A. C. Statistical Mathematics. Fourth Edition. Edinburgh, Scotland: Oliver & Boyd Ltd., 1945. Pp. vi, 153. 5s. (New York: Interscience Publishers, Inc. $1.75.)* [655

American Institute of Electrical Engineers, Subcommittee on Educational Activities (J. Manuele, H. F. Dodge, A. I. Peterson, and R. E. Wareham). Statistical Methods in Quality Control: 9, Statistical Tests of Significant Differences. *Elec Eng* 65(10):466-8 O '46.* [656

Armstrong, G. R., and Clarke, P. C. Frequency Distributiom vs. Acceptance Table. *Ind Qual Control* 3(2):22-7 S '46.* [657

Baker, G. A. Distribution of the Ratio of Sample Range to Sample Standard Deviation for Normal and Combinations of Normal Distributions. *Ann Math Stat* 17(3):366-9 S '46.* [658

Bartlett, Neil R. A Punched-Card Technique for Computing Means, Standard Deviations, and the Product-Moment Correlation Coefficient and for Listing Scattergrams. *Sci* 104 (2703):374-5 O 18 '46.* [659

Baten, W. D., and Trout, G. M. A Critical Study of the Summation-of-Difference-in-Rank Method of Determining Proficiency in Judging Dairy Products. *Biometrics B* 2(4):67-9 Ag '46.* [660

Becker, H. W. Stirling's Numbers of the Third Kind. Abstract. *B Am Math Soc* 52(9):826 S '46.* [661

Birnbaum, Z. W. Tshebysheff Inequality for Two Dimensions. Abstract. *B Am Math Soc* 52(7):621 Jl '46.* [662

Black, Bertram J., and Olds, Edward B. A Punched Card Method for Presenting, Analyzing, and Comparing Many Series of Statistics for Areas. *J Am Stat Assn* 41(235):347-55 S '46.* [663

Blackwell, David. Conditional Expectation and Unbiased Sequential Estimation. Abstract. *B Am Math Soc* 52(9):826 S '46.* [664

Blackwell, D., and Girshick, M. A. On Functions of Sequences of Independent Chance Vectors with Applications to the Problem of the "Random Walk" in *k* Dimensions. *Ann Math Stat* 17(3):310-7 S '46.* [665

Boschan, Paul. The Consolidated Doolittle Technique. Abstract. *B Am Math Soc* 52(9):826 S '46.* [666

Brogden, Hubert E. An Approach

to the Problem of Differential Prediction. *Psychometrika* 11(3):139-54 S '46.* 1667
Brumbaugh, Martin A. Quality Reports for Management. *Elec Eng* 65(8-9):391-3 Ag-S '46.* 1668
Brumbaugh, Martin A. Reporting Quality Results to Management. *Ind Qual Control* 3(2):5-7 S '46.* 1669
Buros, Oscar Krisen. Statistical Methodology Index, No. 4. *J Am Stat Assn* 41(234):270-4 Je '46.* 1670
Buros, Oscar Krisen. Statistical Methodology Index, No. 5. *J Am Stat Assn* 41(235):416-20 S '46.* 1671
Carlton, A. George. Estimating the Parameters of a Rectangular Distribution. *Ann Math Stat* 17(3):355-8 S '46.* 1672
Colley, Russell H. How to Determine the Size of a Survey Sample. *Ptr Ink* 216(10):35+ S 6 '46.* 1673
★Cramér, Harald. Mathematical Methods of Statistics. Uppsala, Sweden: Almqvist & Wiksells, 1945. Pp. xvi, 575. (Princeton, N. J.: Princeton University Press, 1946. $6.00.)* *To be reviewed.* 1674
Croxton, Frederick, and Cowden, Dudley J. Tables to Facilitate Computation of Sampling Limits of *s* and Fiducial Limits of Sigma. *Ind Qual Control* 3(1):18-21 Jl '46.* 1675
Davis, Alfred L. The Quality Control Organization. *Ind Qual Control* (2)3:12-16 S '46.* 1676
De Lury, D. B. The Analysis of Latin Squares When Some Observations Are Missing. *J Am Stat Assn* 41(235):370-89 S '46.* 1677
★Duncan, David B., and Kenney, John F. On the Solution of Normal Equations and Related Topics. Milwaukee 3, Wis.: Book Store, University Extension Division (623 West State St.), 1946. Pp. iii, 35. $1.00. Paper. *Reviewed in this issue.* 1678
Erdős, P., and Kac, M. On Certain Limit Theorems of the Theory of Probability. *B Am Math Soc* 52(4):292-302 Ap '46.* 1679
Esseen, Carl-Gustav. Fourier Analysis of Distribution Functions: A Mathematical Study of the Laplace-Gaussian Law. *Acta Mathematica* 77(1-2):1-125 '45.* 1680
Feller, W. The Law of the Iterated Logarithm for Identically Distributed Random Variables. *Ann Math* 47(4):631-8 O '46.* 1681
Finney, D. J. The Analysis of a Factorial Series of Insecticide Tests. *Ann Appl Biol* 33(2):160-5 My '46.* 1682
Finney, D. J. The Frequency Distribution of Deviates from Means and Regression Lines in Samples from a Multivariate Normal Population. *Ann Math Stat* 17(3):344-9 S '46.* 1683
Finney, D. J. Orthogonal Partitions of the 5×5 Latin Squares. *Ann Eug* 13(1):1-3 Ap '46.* 1684
Finney, D. J. Recent Developments in the Design of Field Experiments: III. Fractional Replication. *J Agric Sci* 36(3):63-8 Jl '46.* 1685
Fisher, R. A. The Fitting of Gene Frequencies to Data on Rhesus Reactions. *Ann Eug* 13(2):150-5 Ag '46.* 1686
Frear, Donald E. H. Punch Cards in Correlation Studies. *Chem & Eng News* 23(22):2077 N 25 '45.* 1687
Ghosh, Manindra Nath. On the Order of Approximation Involved in Laplace's Central Limit Theorem in Probability. *Sankhya* 7(3):323-6 Ap '46.* 1688
Gibian, E. F. Quality Control by Statistical Methods Applied to Line Production. *Ind Qual Control* 3(2):7-12 S '46.* 1689
Girshick, M. A. Contributions to the Theory of Sequential Analysis: II, III. *Ann Math Stat* 17(3):282-98 S '46.* 1690
Goddard, L. S. The Accumulation of Chance Effects and the Gaussian Frequency Distribution. *Philos Mag* 36(257):428-33 Je '45.* 1691
Gaddess, Eugene. Probability and Quality Control. *Am Mach* 90(17):110-1 Ag 15 '46.* 1692
Goodwin, Richard. Innovations and the Irregularity of Economic Cycles. *R Econ Stat* 28(2):95-104 My '46.* 1693
★Grant, Eugene L. Statistical Quality Control. New York 18: McGraw-Hill Book Co., Inc., 1946. Pp. xii, 563. $5.00.* *To be reviewed.* 1694
Grubbs, Frank E. The Difference Control Chart With an Example of Its Use. *Ind Qual Control* 3(1):22-5 Jl '46.* 1695
Guttman, Louis. Enlargement Methods for Computing the Inverse Matrix. *Ann Math Stat* 17(3):336-43 S '46.* 1696
Haldane, J. B. S. The Cumulants of the Distribution of Fisher's 'u_{11}' and 'u_{31}' Scores Used in the Detection and Estimation of Linkage in Man. *Ann Eug* 13(2):122-34 Ag '46.* 1697

Hansen, Morris H.; Hurwitz, William N.; and Gurney, Margaret. Problems and Methods of the Sample Survey of Business. *J Am Stat Assn* 41(234):173–89 Je '46.* [698

Hayes, Samuel P., Jr. Diagrams for Computing Tetrachoric Correlation Coefficients From Percentage Differences. *Psychometrika* 11(3):163–72 S '46.* [699

Henry, John A. Quality Control for the Engineering Student. *Ind Qual Control* 3(2):28–9 S '46.* [700

Hoel, P. G., The Efficiency of the Mean Moving Range. Abstract. *B Am Math Soc* 52(9):827 S '46.* [701

Hollander, W. F. Notes on Graphic Biometric Comparisons of Samples. *Am Nat* 80(793):401–6 Jl-Ag '46.* [702

Holt, C. C., and Samuelson, P. A. The Graphic Depiction of Elasticity of Demand. *J Pol Econ* 54(4):354–7 Ag '46.* [703

Hopkins, J. W. Precision of Assessment of Palatability of Foodstuffs by Laboratory Panels. *Can J Res, Sect F* 24(4):203–14 Jl '46.* [704

Hopkins, J. W. Statistical Design of Experiments. *Can Chem & Process Ind* 30(5):113–6+ My '46.* [705

Hsu, P. L. On the Asymptotic Distributions of Certain Statistics Used in Testing the Independence Between Successive Observations From a Normal Population. *Ann Math Stat* 17(3):350–4 S '46.* [706

Hubback, J. A. Sampling for Rice Yields in Bihar and Orissa. Originally published as Bulletin No. 166 of the Agricultural Research Institute, Pusa. *Sankhyā* 7(3):281–94 Ap '46.* [707

Hughes, H. A. Experimental Data and "Sufficient" Accuracy. Letter. *Nature* 158(4001):29 Jl 6 '46.* [708

Husain, Q. M. Impossibility of the Symmetrical Incomplete Block Design With λ = 2, k = 7. *Sankhyā* 7(3):317–22 Ap '46.* [709

Jones, Howard L. Linear Regression Functions With Neglected Variables. *J Am Stat Assn* 41(235):356–69 S '46.* [710

Juran, Joseph M. The Relation of the Quality Function to the Industrial Enterprise. *Ind Qual Control* 3(1):5–7 Jl '46.* [711

Kac, Mark. A Discussion of the Ehrenfest Model. Abstract. *B Am Math Soc* 52(7):621 Jl '46.* [712

Katz, Leo. On the Class of Functions Defined by the Difference Equation (x+1) f(x+1) = (a+bx) f(x). Abstract. *B Am Math Soc* 52(9):827 S '46.* [713

★Kendall, M. G. Contributions to the Study of Oscillatory Time-Series. National Institute of Economic and Social Research, Occasional Papers, IX. London: Cambridge University Press, 1946. Pp. viii, 76. 7s. 6d. Paper. (New York 11: Macmillan Co. $1.75.) * *To be reviewed.* [714

Kimball, B. F. Some Basic Theorems for Developing Tests of Fit for the Case of the Non-Parametric Probability Distribution Function. Abstract. *B Am Math Soc* 52(9):827 S '46.* [715

Kimball, Bradford F. Sufficient Statistical Estimation Functions for the Parameters of the Distribution of Maximum Values. *Ann Math Stat* 17(3):299–309 S '46.* [716

Knowler, Lloyd A. Fundamentals of Quality Control. *Ind Qual Control* 3(1):7–18 Jl '46.* [717

Laderman, Jack, and Abramowitz, Milton. Application of Machines to Differencing of Tables. *J Am Stat Assn* 41(234):233–7 Je '46.* [718

Lebedeff, George M., and Goode, Henry P. P Charts for Quality Control. *Factory Mgmt* 104(9):129–36 S '46.* [719

Lesser, Arthur, Jr. The Statistical Approach in Industrial Research. *Iron Age* 158(8):50–5 Ag 22 '46.* [720

Levene, Howard. A Test of Randomness in Two Dimensions. *B Am Math Soc* 52(7):621 Jl '46.* [721

Lindley, D. V. Linear "Curves of Best Fit" and Regression Lines. Letter. *Nature* 158(4008):272–3 Ag 24 '46.* [722

Madow, Lillian H. Systematic Sampling and Its Relation to Other Sampling Designs. *J Am Stat Assn* 41 (234):204–17 Je '46.* [723

Mahalanobis, P. C. Sample Surveys of Crop Yields in India. *Sankhyā* 7(3):269–80 Ap '46.* [724

Mathen, K. K. A Criterion for Testing Whether Two Samples Have Come From the Same Population Without Assuming the Nature of the Population. *Sankhyā* 7(3):329 Ap '46.* [725

Molina, Edward C. Some Fundamental Curves for the Solution of Sampling Problems. *Ann Math Stat* 17(3):325–35 S '46.* [726

Mood, A. M. Statistical Control of the Production Process. *Food Ind*

18(8, 9):1180–91, 1372–4+ Ag, S '46.* [727

Mukherjea, Ramkrishna, and Mukherjee, Moni Mohan. A Note on Concentration of Income in Bengal Villages. *Sankhyā* 7(3):327 Ap '46.* [728

Nandi, Hari Kinkar. Enumeration of Non-Isomorphic Solutions of Balanced Incomplete Block Designs. *Sankhyā* 7(3):305–12 Ap '46.* [729

Nandi, Hari Kinkar. A Further Note on Non-Isomorphic Solutions of Incomplete Block Designs. *Sankhyā* 7(3): 313–16 Ap '46.* [730

Panse, V. G. An Application of the Discriminant Function for Selection in Poultry. *J Genetics* 47(3): 242–8 Jl '46.* [731

Parle, H. V. Practical Uses of Probabilities. *Neb Blue Print* 45(1):8–10+ O '45.* [732

Patterson, R. E. The Use of Adjusting Factors in the Analysis of Data With Disproportionate Subclass Numbers. *J Am Stat Assn* 41(235):334–46 S '46.* [733

Penrose, L. S. The Elementary Statistics of Majority Voting. *J Royal Stat Soc* 109(1):53–7 pt 1 '46.* [734

Rao, C. Radhakrishna. Confounded Factorial Designs in Quasi-Latin Squares. *Sankhyā* 7(3):295–304 Ap '46.* [735

Rao, C. Radhakrishna. On the Linear Combination of Observations and the General Theory of Least Squares. *Sankhyā* 7(3):237–56 Ap '46.* [736

Robbins, F. J. Statistical Control Applied to Wrench Manufacture. *Metal Prog* 50(2):287–90 Ag '46.* [737

Rubin, Herman. A Symptotic Distribution of Moments From a System of Linear Stochastic Difference Equations. Abstract. *B Am Math Soc* 52(9): 827 S '46.* [738

Ruist, Erik. Standard Errors of the Tilting Coefficients Used in Confluence Analysis. *Econometrica* 14(3):235–41 Jl '46.* [739

Sandon, Frank. Scores for Ranked Data in School Examination Practice. *Ann Eug* 13(2):118–21 Ag '46.* [740

Satterthwaite, F. E. Retention of Decimal Places in Matrix Calculations. Abstract. *B Am Math Soc* 52(9):828 S '46.* [741

Shafer, J. L. Sampling Incoming Material. *Ind Qual Control* 3(2):16–20 S '46.* [742

Singh, Jagjit. Theories of Probability. *Sankhyā* 7(3):257–62 Ap '46.* [743

Slack, Margaret. The Probability Distributions of Sinusoidal Oscillations Combined in Random Phase. *J Inst Elec Eng, Part 3* 93(22):76–86 Mr '46.* [744

Smallwood, Hugh M. Statistical Methods in Research. *J Chem Ed* 23(7):352–6 Jl '46.* [745

Snedecor, George W., and Haber, E. S. Statistical Methods for an Incomplete Experiment on a Perennial Crop. *Biometrics B* 2(4):61–7 Ag '46.* [746

Stone, Richard. The Analysis of Market Demand. *J Royal Stat Soc* 108(3–4):286–382; discussion, 383–91 '45.* [747

Sukhatme, P. V. Size of Sampling Unit in Yield Surveys. Letter. *Nature* 158(4010):345 S 7 '46.* [748

Szatrowski, Zenon. Calculating the Geometric Mean From a Large Amount of Data. *J Am Stat Assn* 41(234):218–20 Je '46.* [749

Tinbergen, J. Some Measurements of Elasticities of Substitution. *R Econ Stat* 28(3):109–16 Ag '46.* [750

Tukey, J. W. Sampling From Contaminated Distributions: Preliminary Report. Abstract. *B Am Math Soc* 52(9):828 S '46.* [751

Tukey, John W., and Wilks, S. S. Approximation of the Distribution of the Product of Beta Variables by a Single Beta Variable. *Ann Math Stat* 17(3):318–24 S '46.* [752

Verigan, J. F. Presenting Quality Control to Machine Operators. *Ind Qual Control* 3(2):27–8 S '46.* [753

Walsh, John E. On the Power Function of the Sign Test for Slippage of Means. *Ann Math Stat* 17(3):358–62 S '46.* [754

Wernimont, Grant. Use of Control Charts in the Analytical Laboratory. *Ind & Eng Chem, Anal Ed* 18(10):587–92 O '46.* [755

★Wharton, A. S. Quality Through Statistics, Revised Edition. London: Philips Lamps Ltd., 1946.* (Highland Park, N. J.: Gryphon Press, 1946. Pp. iv, 92. $1.75.) *To be reviewed.* [756

Wherry, Robert J., and Gaylord, Richard H. Test Selection With Integral Gross Score Weights. *Psychometrika* 11(3):173–83 S '46.* [757

Wherry, Robert J., and Taylor, Erwin K. The Relation of Multiserial Eta to Other Measures of Correlation.

1946 INDEX OF JOURNALS

1040 INDEX OF AUTHORS

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

VOLUME 41: 1946

NUMBERS 233-236

Published Quarterly by the

AMERICAN STATISTICAL ASSOCIATION

WASHINGTON, D. C.

1946

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

CONTENTS OF VOLUME 41

* Annual Meeting papers are indicated by an asterisk.

ARTICLES

BOOK REVIEWS

BOOK REVIEWERS